

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2002, 11:55:09 ; Search time 281.82 Seconds
(without alignments)
9083.526 Million cell updates/sec

Title: US-09-667-380A-1

Sequence: 1 atgagctgcctcgtggtggtg.....gcatcttctgctcagcag 1491

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_032802.*
1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1491	100.0	1491	22 AAF77686	Human protease-inh
2	1491	100.0	2272	22 AAF77687	Human protease-inh
3	1487.8	99.8	4877	22 AAF77687	Human cancer agent
4	1487.8	99.8	4877	22 AAF77687	Human cancer agent
5	1487.8	99.8	4877	22 AAF77687	Human cancer agent
6	1486.2	99.7	2400	22 AAD17764	Human novel trypsi
7	1474.8	98.9	2403	22 AAD17766	Human novel trypsi
8	1452.6	97.4	2412	22 AAD17767	Human novel trypsi
9	1393.8	93.5	1669	22 AAD17765	Human novel trypsi

10	1286.2	86.3	2305	22 AAD17763	Human novel trypsi
11	1109.6	74.4	1690	22 AAD15690	Human cDNA sequenc
12	724	48.6	939	22 AAD98687	Rat EST-derived co
13	594.6	39.9	840	22 AAD05058	Human cDNA clone (
14	484.4	32.5	3038	22 AAD34199	Human PRO541 nucle
15	484.4	32.5	3038	22 AAD78550	Human PRO541 (UNO3
16	484.4	32.5	3038	22 AAD54959	Human DNA encoding
17	484.4	32.5	3038	22 AAD54959	Human DNA encoding
18	464.8	31.2	2408	22 AAD60849	Human ORF1580
19	440	29.5	561	22 AAD99444	Human proteinenco
20	287.2	19.3	927	22 AAS33053	Human polynucleoti
21	278.2	18.7	1508	22 AAI59063	DNA encoding human
22	274.6	18.4	892	22 AAS33230	Human novel protei
23	274.6	18.4	1498	22 AAD16735	Human novel protei
24	269.4	18.1	1498	22 AAD16738	Human novel protei
25	250.4	16.8	843	22 AAS33345	DNA encoding human
26	250.4	16.8	843	22 AAS33345	Membrane-bound pro
27	240	16.1	1738	21 AAZ65086	Human DNA encoding
28	240	16.1	1738	22 AAS46044	Human PRO1156 (UNQ
29	240	16.1	1738	22 AAF44232	Human PRO1156 (UNQ
30	218.6	14.7	594	16 AAT04077	Trypsin inhibitor
31	218.6	14.7	594	16 AAT04092	Human trypsin inh
32	215.2	14.4	567	16 AAT04085	Human trypsin inh
33	214	14.4	291	22 AAK73276	Human immune/haema
34	203.2	13.6	1520	22 AAS03045	Human diagnostic a
35	202	13.5	301	22 AAK58655	Human immune/haema
36	169.4	11.4	1450	22 AAS31098	Human diagnostic a
37	138	9.3	561	22 AAI10556	Human cDNA clone (
38	126.4	8.5	128	24 ABA90233	Human ORF40 coding
39	83.8	5.6	474	22 AAK76055	Human immune/haema
40	79.8	5.4	1875	20 AAX52263	Human polynucleoti
41	79.8	5.4	1875	20 AAX52263	Protein PRO328 CDN
42	79.8	5.4	1875	21 AAS21423	Human cDNA sequenc
43	79.8	5.4	1875	22 AAF72421	Human PRO328 cDNA
44	79.8	5.4	1875	22 AAF72421	Human angiogenesis
45	79.8	5.4	1875	22 AAC97471	Human angiogenesis

ALIGNMENTS

RESULT 1	AAAF77686	standard; cDNA; 1491 BP.
XX	AAAF77686;	
XX	29-MAY-2001 (first entry)	
XX	Human protease-inhibitor like protein coding sequence #1.	
XX	Human; protease-inhibitor like protein; NHP; novel human protein;	
XX	trypsin inhibitor; gene therapy; drug screening; disease diagnosis;	
XX	signal transduction; ss.	
XX	Homo sapiens.	
XX	WO200121651-A2.	
XX	29-MAR-2001.	
XX	22-SEP-2000; 2000WO-US26048.	
XX	24-SEP-1999; 99US-0156101.	
XX	(LEXI-) LEXICON GENETICS INC.	
XX	Donoho G, Turner CA, Mattler F, Nehls M, Friedrich G;	
XX	Zambrowicz B, Sands AT;	
XX	WPI; 2001-266064/27.	
XX	P-PSDB; AAB74446.	

Db 3397 CCTTCTACTGCGCTCCGGTGTGGACCAACATCTATGATACCTTCAACCATCTGC 3338
 1321 aagaagctgtgacgagagatcaagaagaagtgaggatgacgtgacgtgacg 1380
 3337 AAGACAGCCGTGACGCGGAGATCATGACAGCAGAGTGGGGTACGCGACGTGAG 3278
 1381 cccgtgataaagaagaagctgagtgctgctcagagatgagttcagctgaagc 1440
 3277 CCGGTGATATAAAGAGACCTAGTGGCTCCCTCAGAAATGAGATTGATAAGC 3218
 1441 ctggagactcctcgagatgagagagcttcgagatcttgcgtcagcag 1491
 3217 CTGGGAGACTCTCGGATGGAAGGCGCTTCGGATCTTGTGTCAGCAG 3167

RESULT 5

AAS60883/c
 ID AAS60883 standard; cDNA; 4877 BP.

AC AAS60883;

DT 29-JAN-2002 (first entry)

DE Human cancer agent-resistance marker #542.

KW Human; cancer cell marker; TAXOL; cytostatic; tumour; carcinoma;

KW squamous cell carcinoma; sarcoma; fibrosarcoma; leukaemia;

KW lymphocytic leukaemia; lymphoma; plasmocytoma; reticulum cell sarcoma;

KW Hodgkin's disease; glioma; ss.

OS Homo sapiens.

PN WO200179556-A2.

PD 25-OCT-2001.

PF 13-APR-2001; 2001WO-US12132.

PR 14-APR-2000; 2000US-197538P.

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Lillie J, Brown JL, Bolt A, Van Hufel C;

PS WPI; 2001-602993/66.

PT Novel nucleic acid, used as a marker to determine the effectiveness of
 using TAXOL to treat cancer cell growth in individuals.

Claim 1; Page 382-384; 527pp; English.

The invention relates to 1046 novel nucleic acids which are used as
 markers for determining the sensitivity of a cancer cell to the
 anticancer agent TAXOL. Cancer cells can be treated with TAXOL when
 they are shown to express one of the 242 sensitivity markers or the
 cells are shown not to express one of the 804 resistance markers. The
 methods can be used to determine the effectiveness of TAXOL
 in the treatment of cancer cell growth in an individual. The
 can be used as targets in developing anti-cancer agents such as
 chemotherapeutic compounds. The markers can also be used as targets in
 developing treatments for cancer, particularly those cancers which
 display resistance to agents and exhibit expression of the markers. The
 anticancer agents developed by the novel method can be used to treat
 genomic sequences corresponding to the markers, in the identification of
 cells or tissues which mis-express the protein. Cancers which may
 be targeted include carcinoma (e.g. squamous cell carcinoma),
 sarcoma (e.g. fibrosarcoma), leukaemia (e.g. lymphocytic leukaemia),
 lymphoma, plasmocytoma, reticulum cell sarcoma, Hodgkin's disease and
 tumours (e.g. glioma). The present sequence is one of the 1046
 novel cancer cell markers.

SO Sequence 4877 BP; 1183 A; 1201 C; 1236 G; 1257 T; 0 other;

Query Match 99.8%; Score 1487.8; DB 22; Length 4877;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1489; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgagctgcgttcctggtgtgtgtcaacccttggtgctgtctgtctgtcgtgac 60
 Db 4657 ATGAGCTGCGTCTCGGTGTGTGTCAACCTTGGGGTGTCTTCTCGGTGTGGATCC 4598
 QY 61 caaggctaccctctgcccgaagcttctctttagagagctgtcagcaatacagcac 120
 Db 4597 CAAGGCTACCTCTCTGCCCCAACGTCCTCTTAGAGAGAGCTGTCAAGAAATACCGAC 4538
 QY 121 aacgagctcaccctccgggttcgagagccatcccaaggagagagagagctctc 180
 Db 4537 AACGAGCTCACTCCCGGTTCCGACAGCCATCCAGGAGAGACAAAGAGAGATCCTC 4478
 QY 181 atgtgcacaaagagcttcggggccaggtgtgagctcaggtctcaaatgagatc 240
 Db 4477 ATGCTGCACAAAGCTTCCGGGSCAGGTGCACCTCAAGCCTCCAACTGGAGTACATG 4418
 QY 241 accctggatgaagactgtgagagctgtcagcgtgtggccagtcagtcagtcag 300
 Db 4417 ACCTGGGATGACGAACGTGAGAGTCTGCTCAACCTGGGCCAGTCACTGATCTGGGAG 4358
 QY 301 caaggcccaacagctgtcgtgtgtcatctcgggcaaaccttggcgtcacttggggag 360
 Db 4357 CAGGCGCCACACGTCGCTGCTGTGTCATCGGCGAGAACCTGGCGGCTCACTGGGGCAGG 4298
 QY 361 tctcgtctccgggtgttcacatgtcagtcctgtgtatgcagagtgagagactac 420
 Db 4297 TATCGCTCCGGGTTCCATGTCAGTCTGCTGTGACAGGTGAAGACTACCTTAC 4238
 QY 421 cctaccggagcagatggaacccctgtgtccagagagtgctcgggacctagtgcag 480
 Db 4237 CCTTACCCGAGGAGTGCACCCCTGTGTCACAGAGGTGTCGGGGCTATGTCACAG 4178
 QY 481 caactacacagatagttgtggccacacacaaagatggtgtgtctgtgaacactgc 540
 Db 4177 CACTACACACAGATAGTTGGGCCACACACAAAGTCTGCTGTGTAACACCTGC 4118
 QY 541 cgaagatgactgtctgtggagagatgttgaggaacggctcacttctgcacattac 600
 Db 4117 CGGAGATGACTGTCTGGGAGAGATTGGAGAACGGCTCTATTGTCGAATTAT 4058
 QY 601 tctccaaagggagctgatatgagagagccctacaaagatggccgctgtctgcag 660
 Db 4057 TCTCCAAAGGGGAACTGATGGAAGACCCCTACAAAGATGGCCCGCTGCTCTGAG 3998
 QY 661 tggccaccagctatgagagcaactgcaggaacaacttgtttacaggaagaacacac 720
 Db 3997 TGCCACCCAGACTATGAGAGCAGCTCCAGAGAACACTTGTGTACCGAAGAAACCTAC 3938
 QY 721 actccaacactgaaagcagatgaaatgagtgagtgaggaagcgtccactctgaagaa 780
 Db 3937 ACTCCAAACCTAAAGCGAGATGATGAGTGAAGCGGCTCCATTCTCGAAGAA 3878
 QY 781 aaccattgtgtgtcacaacggaggtgatgagaccacaaagcccaagaacactctg 840
 Db 3877 AACCATTTTGGCTCCAAACCGAGGGGTGATGACACCCCAAGCCCAAGAAACCTCTGCG 3818
 QY 841 gtaactacatgaccacaaatcgtcagatgttacaccagaatgaaagcaggtgcaag 900
 Db 3817 GTCAACTACATGACCAAGTGTCAAGATGTACACCAAGATGAAGGACAGCTCAAGAGG 3758
 QY 901 tcaagtgttaagagttaccagttgccagcaggtcgtcgaacccaagaggaagacttt 960
 Db 3757 TCACAGTGTAAACAGGTACAGAGTGCACAGCTGCTGAACCAAGAGGAGATCTTT 3698
 QY 961 ggaagctgtttatgaagactgctcagagtatgcccgcgcacatcaactagggatc 1020

Db	3697	GGAACTCTGTTTCATGAAACCTGTCACATATGCCGGCCGCATCACTACAGGATC	3638
QY	1021	ctggaatacaagaaggagccggtgagatatacaacagaagaaggaaagtcctctctctgtg	1080
Db	3637	CTGAGATACAAAGGAGGCGCTGGTGGATATCCAGAGAAAGGAAAGTCCCTCTCTCGTG	3578
QY	1081	aagctctagagacacgycgtgcagtcctccctcaagaatacaaacctctcagctcatatcgt	1140
Db	3577	AAGTCTAGAGACACGGCGCTGAGTCCCTCAGCAAAATCAAACTTCCAGCTCATTTATG	3518
QY	1141	gtgtcacaagtgaagtgtcagaatttggacctctacaegaaccgttgtcagctgtgccg	1200
Db	3517	GTGTCAAAATGTGAATGCAAGGATTTTGGACTCTACACACCCTGTCTACACTGTGCCG	3458
QY	1201	tttbaaagcacaacaactcactccccaagaatccaattgtccgagcaactgtgaagaagaa	1260
Db	3457	TTTGAAGAGCCACCACTCACTCCCAAAATCACTTGTCCGGCACACTGCAAAAGACGA	3398
QY	1261	ccttctactctgggctcccgatgttttgaaaccaacatctatgacatacctcaagaatctgc	1320
Db	3397	CCCTTCACACGGGCTCCGGTGTGGAAACCAATCTATGCAAGATACCTCAAGCATCTGC	3338
QY	1321	aagacacgctgtgacgcgagagatcatcaacaacagagatgtgggggtggaagtgagacgtatg	1380
Db	3337	AAGACACCCCTGCACCGGAGATCATACCAACGAGAGTGGGGGTGAAGTGGAGCGTATG	3278
QY	1381	cccgctgataaaaaagaagaaactcagctgagctcgtcctcaggaatgtagttcagctcgaagc	1440
Db	3277	CCCGTGATTTAAAAGAAAGACTACGCGGGCTGCTCAGAGAAAGAGTTCACTGTGAAGC	3218
QY	1441	cttggagacccctctggagatggaagagccttcgagatcttgcctgtcaagcag	1491
Db	3217	CTGGGACATCTCTGGGATGCAAAAGGCGCTCCCGGATCTTGTGCTGATGGCAG	3167

RESULT	6
AAD17764	
ID	AAD17764 standard; DNA; 2400 BP.
AC	
XX	AAD17764;
DT	10-DEC-2001 (first entry)
XX	
DE	Human novel trypsin inhibitor-like protein, NOV-4b encoding DNA.
KW	Human; NOV-X protein; KIAA1233-like protein; STE20-like protein; tumour
KW	trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
KW	immunological disorder; neurodegenerative disorder; Alzheimer's disease
KW	Parkinson's disease; immunomodulatory; pharmacogenomic; hemostatic;
KM	human immunodeficiency virus; HIV; fertility disorder; neutroprotective;
KM	cytostatic; neotropic; anti-fertility; cancer; NOV-4b protein; ds.
OS	
XX	Homo sapiens.
FH	Key
FT	CDS
XX	
FT	Location/Qualifiers
FT	206..1699
FT	/tag= a
FT	/product= "Human novel trypsin inhibitor-like protein,
FT	NOV-4b"
FT	206..271
FT	/tag= b
FT	272..1696
FT	/tag= c
FT	/product= "Human mature novel trypsin inhibitor-like
FT	protein, NOV-4b"
PN	WO200162928-A2.
XX	
PD	30-AUG-2001.
XX	
PF	26-FEB-2001; 2001MO-US06151.
XX	
PR	25-FEB-2000; 2000US-0184951.

PR	28-FEB-2000;	2000US-0185548.
PR	01-MAR-2000;	2000US-0185567.
PR	18-APR-2000;	2000US-0197723.
PR	27-APR-2000;	2000US-0199957.
PR	23-FEB-2001;	2001US-0789390.
XX		
PA	(CURA-) CURAGEN CORP.	
PI	Vennet CAM, Fernandes E, Shinkets RA, Macdougall J, Spaderna SK:	
DR	WPI; 2001-582051/55.	
DR	P-PSDB; AAEI0616.	
PT	New isolated KIAI1233-like, SRE20-like, or trypsin inhibitor-like	
PT	polypeptide for diagnosing and treating pathological disorders, such as	
XX	Parkinson's disease and for use in pharmacogenomics -	
XX		
PS	Claim 9; Page 83; 189pp; English.	
CC	The invention relates to novel human polypeptides referred as NOV-X	
CC	and their corresponding nucleic acid sequences. NOV-X collectively	
CC	include NOV-1, NOV-2a and NOV-2b which are novel KIAI1233-like	
CC	polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel	
CC	SRE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e	
CC	which are novel trypsin inhibitor-like polypeptides. NOV-X is used to	
CC	identify a potential therapeutic agent that can modulate its activity	
CC	and can be used for treating a pathology related to aberrant expression	
CC	or aberrant physiological interactions of NOV-X. NOV-X or its DNA is	
CC	used to determine the presence or predisposition to a disease associated	
CC	with altered levels of NOV-X. NOV-X, its DNA and its antibody are used	
CC	to treat or prevent a pathology associated with NOV-X. The pathological	
CC	states that can be treated or prevented are haematopoietic, cancer,	
CC	immunological, tumour, neurodegenerative (e.g., Alzheimer's and	
CC	Parkinson's disease), human immunodeficiency virus (HIV) illness and	
CC	fertility disorders. NOV-X and its DNA are used in pharmacogenomics for	
CC	predictive medicine. NOV-X DNA is used in gene therapy. The present	
CC	sequence is a DNA encoding human novel trypsin inhibitor-like protein,	
XX	NOV-4b.	
SQ	Sequence 2400 BP; 556 A; 646 C; 684 G; 514 T; 0 other;	
Query Match	99.7%; Score 1486.2; DB 22; Length 2400;	
Best Local Similarity	99.8%; Pred. No. 0;	
Matches 1488; Conservative	0; Mismatches 3; Indels 0; Gaps	0;
OY	1 atagactgcgctctgggttgatctaccccttggggcgcttcctcgctcggatcc 60	
Db		
OY	206 atgaactgcgctctgggttgatctaccccttggggcgcttcctcgctcggatcc 265	
OY	61 caagactcactctcgcccacgtaactctctttaaggagaagtctcaagacaataaccagac 120	
Db		
OY	266 caagactcactctcgcccacgtaactctctttaaggagaagtctcaagacaataaccagac 325	
OY	121 aaagaattcactccccgggttcgcgaagccatcccccaggagagaaaggagagatccctc 180	
Db		
OY	326 aaagaattcactccccgggttcgcgaagccatcccccaggagagaaaggagagatccctc 385	
OY	181 atcctcacacaagaagcttcggggcagaatgtagacctaaagctccaacatgagatatcatg 240	
Db		
OY	386 atcctcacacaagaagcttcggggcagaatgtagacctaaagctccaacatgagatatcatg 445	
OY	241 acctggatgacgaactgtagaagctctgctcgaagcgttgggccagttcagttcattctggag 300	
Db		
OY	446 acctggatgacgaactgtagaagctctgctcgaagcgttgggccagttcagttcattctggag 505	
OY	301 caagggcccaacagcttgtgtgtctcaatgggagaagacctggggcgctcactctgggcag 360	
Db		
OY	506 caagggcccaacagcttgtgtgtctcaatgggagaagacctggggcgctcactctgggcag 565	
OY	361 tatgcctctccgggttccatgtgagctctcgtgtalacagagtgtaaggactaacctac 420	
Db		
OY	566 tatgcctctccgggttccatgtgagctctcgtgtalacagagtgtaaggactaacctac 625	

PR 01-MAR-2000; 2000US-0185967.
PR 18-APR-2000; 2000US-0197723.
PR 27-APR-2000; 2000US-0199957.
PR 23-FEB-2001; 2001US-0789390.

XX (CURA-) CUNAGEN CORP.

PI Vernet CAM, Fernandes E, Shinkels RA, Macdougall J, Spaderma SK;

DR WPI; 2001-582051/65.

XX P-PSDB; AAE10619.

PT New isolated KIAA1233-like, STE20-like, or trypsin inhibitor-like
PT polypeptide for diagnosing and treating pathological disorders, such as
PT Parkinson's disease and for use in pharmacogenomics -

XX Claim 9; Page 90-91; 189pp; English.

CC The invention relates to novel human polypeptides referred as NOV-X
CC and their corresponding nucleic acid sequences. NOV-X collectively
CC include NOV-1, NOV-2a and NOV-2b which are novel KIAA1233-like
CC polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel
CC STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e
CC which are novel trypsin inhibitor-like polypeptides. NOV-X is used to
CC identify a potential therapeutic agent that can modulate its activity
CC and can be used for treating a pathology related to aberrant expression
CC or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
CC used to determine the presence or predisposition to a disease associated
CC with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
CC to treat or prevent a pathology associated with NOV-X. The pathological
CC states that can be treated or prevented are hematopoietic, cancer,
CC immunological, tumour, neurodegenerative (e.g. Alzheimer's and
CC Parkinson's disease), human immunodeficiency virus (HIV) illness and
CC fertility disorders. NOV-X and its DNA are used in pharmacogenomics for
CC predictive medicine. NOV-X DNA is used in gene therapy. The present
CC sequence is a DNA encoding human novel trypsin inhibitor-like protein.
CC NOV-4e.

CC Sequence 2412 BP; 560 A; 649 C; 688 G; 515 T; 0 other;

Query Match 97.4%; Score 1452.6; DB 22; Length 2412;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1487; Conservative 0; Mismatches 4; Indels 12; Gaps 2;

QY 1 atgagctgcgtccgtggtgtgtcaatcccttgggtgctgtctgtctgtcgtgac 60
DB |||||||
QY 206 atgagctgcgtccgtggtgtgtcaatcccttgggtgctgtctgtctgtcgtgac 265
DB |||||||
QY 61 caaggtcactctcgtcccaagcttactctcttaagagagctgtcctcagcaataccagac 120
DB |||||||
QY 266 caaggtcactctcgtcccaagcttactctcttaagagagctgtcctcagcaataccagac 325
DB |||||||
QY 121 aacgagctcactccgtggtcgcagagccatccccaaggagagagagagatctc 180
DB |||||||
QY 326 aacgagctcactccgtggtcgcagagccatccccaaggagagagagatctc 385
DB |||||||
QY 181 atgtgtcacaagaagcttcggggtcaggtgcagctcaggtccacaatagtagatc 240
DB |||||||
QY 386 atgtgtcacaagaagcttcggggtcaggtgcagctcaggtccacaatagtagatc 445
DB |||||||
QY 241 accctggagatgacgaactcgaagaagcttcgtctcagcgttcggccagctcagctcgtgag 300
DB |||||||
QY 446 accctggagatgacgaactcgaagaagcttcgtctcagcgttcggccagctcagctcgtgag 505
DB |||||||
QY 301 caagggcccaacgaactctgtctgtgttccatcgggagagagacgttcgctcgtgagag 360
DB |||||||
QY 506 caagggcccaacgaactctgtctgtgttccatcgggagagagacgttcgctcgtgagag 565
DB |||||||
QY 361 tatgtctctccgggttccatcgtgtcagctcctgtgtatgaagagtgaaagatacactac 420
DB |||||||
QY 566 tatgtctctccgggttccatcgtgtcagctcctgtgtatgaagagtgaaagatacactac 625
DB |||||||

QY 421 ccttaaccgagcagatgcaaccctgtgtccagaagaagtgctcgggctatgtgacg 480
DB |||||||
QY 626 ccttaaccgagcagatgcaaccctgtgtccagaagaagtgctcgggctatgtgacg 605
DB |||||||
QY 481 cactacaca-----cagatgttggcccaacaaagaatcgtgtgtgctgtg 531
DB |||||||
QY 606 cactacacaagtaactcagatagtttggcccaacaaagaatcgtgtgtgctgtg 745
DB |||||||
QY 532 aacacgtccgggaagatgactgtctgtgggaagatttgggaagacgcgtctctgtc 591
DB |||||||
QY 746 aacacgtccgggaagatgactgtctgtgggaagatttgggaagacgcgtctctgtc 805
DB |||||||
QY 592 tgcattatctccaaa---gggaactgtgattggagaagccctcacaagaatgtgcgg 648
DB |||||||
QY 806 tgcattatctccaaaagagaggtgattgtggaagccctcacaagaatgtgcgg 865
DB |||||||
QY 649 cctgtctctgagtgcccaaccagatctggaagcagctgtcaggaacaactgtgttcag 708
DB |||||||
QY 866 cctgtctctgagtgcccaaccagatctggaagcagctgtcaggaacaactgtgttcag 925
DB |||||||
QY 709 gaagaacactaacctcacaacaaactgaacagcagagatgaatgaagtgaacgcgtcc 768
DB |||||||
QY 926 gaagaacactaacctcacaacaaactgaacagcagagatgaatgaagtgaacgcgtcc 905
DB |||||||
QY 769 attctgaagaacaaactgtgtgtgtcacaacagaggtgtgtgagaccacaagcccaag 828
DB |||||||
QY 986 attctgaagaacaaactgtgtgtgtcacaacagaggtgtgtgagaccacaagcccaag 1045
DB |||||||
QY 829 aaaacctctcgtgtcaactacatgacccaagctgtgaagtgtgacccaagaatgaaagac 888
DB |||||||
QY 1046 aaaacctctcgtgtcaactacatgacccaagctgtgaagtgtgacccaagaatgaaagac 1105
DB |||||||
QY 889 aggtgcgaaggtgcacagtggtacaggtaccaggtgcacaggtcgtctgtaaccacaag 948
DB |||||||
QY 1106 aggtgcgaaggtgcacagtggtacaggtaccaggtgcacaggtcgtctgtaaccacaag 1165
DB |||||||
QY 949 gcgaagaacttgggaagctgtctctatgaagaagctgtctatgaatgcgcgcgcacac 1008
DB |||||||
QY 1166 gcgaagaacttgggaagctgtctctatgaagaagctgtctatgaatgcgcgcgcacac 1225
DB |||||||
QY 1009 cactacggagctcctgtgatacgaaggagcgtgtgtgatacacaagaaggagagtc 1068
DB |||||||
QY 1226 cactacggagctcctgtgatacgaaggagcgtgtgtgatacacaagaaggagagtc 1285
DB |||||||
QY 1069 cccttctctgtgaagctgtgagagacacggcgtgtcaatccctcagaacaatacaaaccttcc 1128
DB |||||||
QY 1286 cccttctctgtgaagctgtgagagacacggcgtgtcaatccctcagaacaatacaaaccttcc 1345
DB |||||||
QY 1129 agctcaatcaatgtgtcaaaagtgaagtgcaggaatttggactgtacacgacgttgc 1188
DB |||||||
QY 1346 agctcaatcaatgtgtcaaaagtgaagtgcaggaatttggactgtacacgacgttgc 1405
DB |||||||
QY 1189 cagctgtgcctgttgaagaagcagaactcaatctgtcccaagaatcattgtccggcaac 1248
DB |||||||
QY 1406 cagctgtgcctgttgaagaagcagaactcaatctgtcccaagaatcattgtccggcaac 1465
DB |||||||
QY 1249 tgcagaagcagaacactcctcactgtggtcgtgtgttggagaacaacatcctacagatacc 1308
DB |||||||
QY 1466 tgcagaagcagaacactcctcactgtggtcgtgtgttggagaacaacatcctacagatacc 1525
DB |||||||
QY 1309 tcaagatctgcagaagcagctgtgacgcgggaatcacaagaacagagatgggggtgac 1368
DB |||||||
QY 1526 tcaagatctgcagaagcagctgtgacgcgggaatcacaagaacagagatgggggtgac 1585
DB |||||||
QY 1369 gtgagcgttatgtccgtgtgataaagaagacactcgtgtgtgtcgtcagaatgtgatt 1428
DB |||||||
QY 1586 gtgagcgttatgtccgtgtgataaagaagacactcgtgtgtgtcgtcagaatgtgatt 1645
DB |||||||
QY 1429 cagctgcgaagccttgggagctcctcgtgagtggaagagccttcggaatcttgcgtcagg 1488
DB |||||||
QY 1646 cagctgcgaagccttgggagctcctcgtgagtggaagagccttcggaatcttgcgtcagg 1705
DB |||||||
QY 1489 cag 1491

Df	1706 cag 1708
RESULT	9
ID	AAD17765
AC	AAD17765 standard; DNA; 1669 BP.
AA	AAD17765;
DE	10-DEC-2001 (first entry)
XX	Human novel trypsin inhibitor-like protein, NOV-4c encoding DNA.
XX	Human: NOV-X protein; KIAI1233-like protein; STE20-like protein; tumour;
KW	trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
KW	immunological disorder; neurodegenerative disorder; Alzheimer's disease;
KW	Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
KX	human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
XX	cytostatic; nootropic; anti-fertility; cancer; NOV-4c protein; ds.
OS	Homo sapiens.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	205..1638
FT	/tag= a
FT	/product= "Human novel trypsin inhibitor-like protein,
FT	NOV-4c"
FT	sig_peptide 205..270
FT	/tag= b
FT	mat_peptide 271..1635
FT	/tag= c
FT	/product= "Human mature novel trypsin inhibitor-like
XX	protein, NOV-4c"
PN	WO200162928-A2.
PD	30-AUG-2001.
PF	26-FEB-2001; 2001WO-US06151.
PR	25-FEB-2000; 2000US-0184951.
PR	28-FEB-2000; 2000US-0185548.
PR	01-MAR-2000; 2000US-0185967.
PR	18-APR-2000; 2000US-0197723.
PR	27-APR-2000; 2000US-0199957.
PR	23-FEB-2001; 2001US-0789390.
PA	(CURA-) CURAGEN CORP.
PI	Vernet CAM, Fernandes E, Shimkets RA, Macdougall J, Spaderna SK;
DR	WPJ; 2001-582051/65.
DR	P-PDB; AAEL10617.
PT	New isolated KIAI1233-like, STE20-like, or trypsin inhibitor-like
PT	polypeptide for diagnosing and treating pathological disorders, such as
PT	Parkinson's disease and for use in pharmacogenomics -
PS	Claim 9; Page 85; 189pp; English.
XX	The invention relates to novel human polypeptides referred as NOV-X
CC	and their corresponding nucleic acid sequences. NOV-X collectively
CC	include NOV-1, NOV-2a and NOV-2b which are novel KIAI1233-like
CC	polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel
CC	STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e
CC	which are novel trypsin inhibitor-like polypeptides. NOV-X is used to
CC	identify a potential therapeutic agent that can modulate its activity
CC	and can be used for treating a pathology related to aberrant expression
CC	or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
CC	used to determine the presence or predisposition to a disease associated
CC	with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
CC	to treat or prevent a pathology associated with NOV-X. The pathological

Query Match	Best Local Similarity	93.5%	Score 1393.8	DB 22	Length 1669
Matches 1401	Conservative	99.2%	Pre. No. 0	Mismatches 12	Indels 0
					Gaps 0
OY 1	atgagctgcgtccctgggtggtgtgatactccctctggggctgctgtctctgctgagctcc	60			
Db 205	atgagctgcgtccctgggtggtgtgatactccctctggggctgctgtctctgctgagctcc	264			
OY 61	caaggctacctctcgtcccaacgctacctctctatagaagagctgtctcaagcaaatacagcaac	120			
Db 265	caaggctacctctcgtcccaacgctacctctctatagaagagctgtctcaagcaaatacagcaac	324			
OY 121	aacgagcttcaactcccggttcgcgagagccatcccccagaggaagaaagagagatctctc	180			
Db 325	aacgagcttcaactcccggttcgcgagagccatcccccagaggaagaaagagagatctctc	384			
OY 181	atgcgtgcacaacaagctcttcgggcagagtgacagctcagcctccacaatagagatactg	240			
Db 385	atgcgtgcacaacaagctcttcgggcagagtgacagctcagcctccacaatagagatactg	444			
OY 241	acctgggatgatgcgaacttgagagaatctctgtcagcgttgaggccatgcattgtcctgggag	300			
Db 445	acctgggatgatgcgaacttgagagaatctctgtcagcgttgaggccatgcattgtcctgggag	504			
OY 301	caagggcccaacagctctgtctgttccatctcgggcagagaaactctgggcgtctacatcgggacag	360			
Db 505	caagggcccaacagctctgtctgttccatctcgggcagagaaactctgggcgtctacatcgggacag	564			
OY 361	tatgcgtctcggggttcccatgtgcagctccgtgtatgacagagtcgaagatacaactac	420			
Db 565	tatgcgtctcggggttcccatgtgcagctccgtgtgtatgacagagtcgaagatacaactac	624			
OY 421	cccttaccgagagagtgacacccctctgtgttccagagaagtgcttcggggctatgtgcag	480			
Db 625	cccttaccgagagagtgacacccctctgtgttccagagaagtgcttcggggctatgtgcag	684			
OY 481	caactacacacaatagtttgggcgcacacacaacaagaatcgtgtctgttgcgttgacaacttgc	540			
Db 685	caactacacacaatagtttgggcgcacacacaacaagaatcgtgtgttgcgttgacaacttgc	744			
OY 541	cggagaaatgactgctctggggagagaagtgttggggagaacgcgtctacttctctgcaattat	600			
Db 745	cggagaaatgactgctctggggagagaagtgttggggagaacgcgtctacttctctgcaattat	804			
OY 601	tctccaaaggggaactggtcttgagagaagcccccatacaagaatctggccggccctgcgtctgag	660			
Db 805	tctccaaaggggaactggtcttgagagaagcccccatacaagaatctggccggccctgcgtctgag	864			
OY 661	tgcgccaccagctatgagaggaagctgtgcagaagaacaacttgttataccgagaagaacaactac	720			
Db 865	tgcgccaccagctatgagaggaagctgtgcagaagaacaacttgttataccgagaagaacaactac	924			
OY 721	actccaaaaccttgaaaacgagacgagatgaaatgagtgagtgaaaacggtcccatccttgaaaga	780			
Db 925	actccaaaaccttgaaaacgagacgagatgaaatgagtgagtgaaaacggtcccatccttgaaaga	984			
OY 781	aaccatgttttggtcccaacgcgaggtgtgatgagacccacaagcccaagaagaacctctgcg	840			
Db 985	aaccatgttttggtcccaacgcgaggtgtgatgagacccacaagcccaagaagaacctctgcg	1044			
OY 841	gttaactaatatgacccaagtctgcgaatgtgatacgaagaatgagaacaggtgtgcaaaagg	900			

Db 1045 gtcaactaacatgaccgaagtcgtctatgtagacacaaatgaaagagagagag 1104
 QY 901 tccacgtgtatcacagatgacagatgacagagagagagagagagagagagag 960
 Db 1105 tccacgtgtatcacagatgacagatgacagagagagagagagagagagagag 1164
 QY 961 ggaagtcgtctatgtagacatgtagacatgtagacatgtagacatgtagac 1020
 Db 1165 ggaagtcgtctatgtagacatgtagacatgtagacatgtagacatgtagac 1224
 QY 1021 ctgagatgacag 1080
 Db 1225 ctgagatgacag 1284
 QY 1081 aagtcgtgagacag 1140
 Db 1285 aagtcgtgagacag 1344
 QY 1141 ggtgcaaaagtgaaatgcaagatgtagacatgtagacatgtagacatgtagac 1200
 Db 1345 ggtgcaaaagtgaaatgcaagatgtagacatgtagacatgtagacatgtagac 1404
 QY 1201 ttggaagaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 1260
 Db 1405 ttggaagaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 1464
 QY 1261 ccttcctactggtgctcggtgttggaaacaaatctatgacatgacatgacatgac 1320
 Db 1465 ccttcctactggtgctcggtgttggaaacaaatctatgacatgacatgacatgac 1524
 QY 1321 aagacagcgtgacagcggagatctacagcaagagagagagagagagagagagag 1380
 Db 1525 aagacagcgtgacagcggagatctacagcaagagagagagagagagagagagag 1584
 QY 1381 cccgtgagataaaagaagacacacacacacacacacacacacacacacacac 1443
 Db 1585 cccgtgagataaaagaagacacacacacacacacacacacacacacacacac 1617

RESULT 10

AAD17763
 ID AAD17763 standard; DNA; 2305 BP.

AC AAD17763;

DT 10-DEC-2001 (first entry)

DE Human novel trypsin inhibitor-like protein, NOV-4a encoding DNA.

XX Human: NOV-X protein: KIAA1233-like protein; STE20-like protein; tumour;
 KM trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
 KM immunological disorder; neurodegenerative disorder; Alzheimer's disease;
 KM Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
 KM human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
 KM cytosolic; neotropic; anti-infectivity; cancer; NOV-4a protein; ds.

OS Homo sapiens.

FH Key Location/Qualifiers
 FT CDS 453..1604
 FT /product= a
 FT /product= "Human novel trypsin inhibitor-like protein,
 FM NOV-4a"

MO200162928-A2.

30-AUG-2001.

26-FEB-2001; 2001WO-US06151.

XX 25-FEB-2000; 2000US-0184951.
 PR 28-FEB-2000; 2000US-018548.
 PR 01-MAR-2000; 2000US-0185967.

PR 18-APR-2000; 2000US-0197723.
 PR 27-APR-2000; 2000US-0199957.
 PR 23-FEB-2001; 2001US-0789390.
 PA (CURA-) CURAGEN CORP.
 PI Vernet CAM, Fernandes E, Shinkets RA, Macdougall J, Spaderna SK;
 DR WPI: 2001-582051/65.
 P-PSDB; AAE10615.
 PT New isolated KIAA1233-like, STE20-like, or trypsin inhibitor-like
 PT polypeptide for diagnosing and treating pathological disorders, such as
 PT Parkinson's disease and for use in pharmacogenomics -
 PS Claim 9; Page 80-81; 189pp; English.
 XX The invention relates to novel human polypeptides referred as NOV-X
 CC and their corresponding nucleic acid sequences. NOV-X collectively
 CC include NOV-1, NOV-2a and NOV-2b which are novel KIAA1233-like
 CC polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel
 CC STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e
 CC which are novel trypsin inhibitor-like polypeptides. NOV-X is used to
 CC identify a potential therapeutic agent that can modulate its activity
 CC and can be used for treating a pathology related to aberrant expression
 CC or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
 CC used to determine the presence or predisposition to a disease associated
 CC with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
 CC to treat or prevent a pathology associated with NOV-X. The pathological
 CC states that can be treated or prevented are haematopoietic, cancer,
 CC immunological, tumour, neurodegenerative (e.g. Alzheimer's and
 CC Parkinson's disease), human immunodeficiency virus (HIV) illness and
 CC fertility disorders. NOV-X and its DNA are used in pharmacogenomics for
 CC predictive medicine. NOV-X DNA is used in gene therapy. The present
 CC sequence is a DNA encoding human novel trypsin inhibitor-like protein,
 CC NOV-4a.
 XX
 SQ Sequence 2305 BP; 542 A; 619 C; 647 G; 497 T; 0 other;

Query Match 86.3%; Score 1286.2; DB 22; Length 2305;
 Best Local Similarity 93.4%; Pred. No. 0;
 Matches 1393; Conservative 0; Mismatches 3; Indels 95; Gaps 1;

QY 1 atgagctgcgtcgtggtgtgcatccctgtgagctgctgtctgtctgtgagatcc 60
 Db 206 atgagctgcgtcgtggtgtgcatccctgtgagctgctgtctgtctgtgagatcc 265
 QY 61 caaggtactctctgccaagctactctcttagagagctgtcagcaaatcacagcac 120
 Db 266 caaggtactctctgccaagctactctcttagagagctgtcagcaaatcacagcac 325
 QY 121 aagagcttactcccggtccgacagacatcccaaggagagagagagatctc 180
 Db 326 aagagcttactcccggtccgacagacatcccaaggagagagagagatctc 385
 QY 181 atgtgcacacaagctcgggagcaggtgacagcctcagagcctcaactgagatcatg 240
 Db 386 atgtgcacacaagctcgggagcaggtgacagcctcagagcctcaactgagatcatg 445
 QY 241 acctggagatgacgaactggaagatctgtcagcgltggccagtcagtcctcggag 300
 Db 446 acctggagatgacgaactggaagatctgtcagcgltggccagtcagtcctcggag 459
 QY 301 cactggccacacagctgtgtgtgtcactcggcagagacactggcgctcaactgggacag 360
 Db 460 cactggccacacagctgtgtgtgtcactcggcagagacactggcgctcaactgggacag 470
 QY 361 tatcgctctcggggttccatgtgacgtctgtgtatgagagagtgaaagactaacctac 420
 Db 471 tatcgctctcggggttccatgtgacgtctgtgtatgagagagtgaaagactaacctac 530
 QY 421 ccctaccagcagatgcaacccctgtgttccagagaggtgtctcggggtcctatgtgacg 480

```

Db 531 ccttaccgagcgagtgcaaccccttggttccagagaggtgctcggggactatgtgcag 590
Qy 481 cactacacacagatagttgggcccaccacaacaagaatggttgcgttgtaacacctgc 540
Db 591 cactacacacagatagttgggcccaccacaacaagaatggttgcgttgtaacacctgc 650
Qy 541 cggagatagctgtctggggagagatttggagaaagcggtcacttctcgaacttat 600
Db 651 cggagatagctgtctggggagagatttggagaaagcggtcacttctcgaacttat 710
Qy 601 tctccaaaggagactgatttgagagaagcccttacaagaatggtcgccgtctctgag 660
Db 711 tctccaaaggagactgatttgagagaagcccttacaagaatggtcgccgtctctgag 770
Qy 661 tgcaccacagctatgagagcagctgcagagaaacttgttaccgagagaagaacctac 720
Db 771 tgcaccacagctatgagagcagctgcagagaaacttgttaccgagagaagaacctac 830
Qy 721 actccaaaacttgaaaagcgagatgaatgaggttgaaaacggtcccttcttgaaga 780
Db 831 actccaaaacttgaaaagcgagatgaatgaggttgaaaacggtcccttcttgaaga 890
Qy 781 aacctgtttggtcccaacgaggtgtagatgagaccacaagcccaagaanaacctctgc 840
Db 891 aacctgtttggtcccaacgaggtgtagatgagaccacaagcccaagaanaacctctgc 950
Qy 841 gtaacatacatgacccaagtctgcaatgtgtgacacccaagttagaagcaggtcagaag 900
Db 951 gtaacatacatgacccaagtctgcaatgtgtgacacccaagttagaagcaggtcagaag 1010
Qy 901 tccacgtgtacacaggtacacacagctgcccagcagctgcttgcagcaacgaagsgaagatctt 960
Db 1011 tccacgtgtacacaggtacacacagctgcccagcagctgcttgcagcaacgaagsgaagatctt 1070
Qy 961 ggaagctgttctatgaaagctcgtctagatgacatgacggtcgccgacatcactagagatc 1020
Db 1071 ggaagctgttctatgaaagctcgtctagatgacatgacggtcgccgacatcactagagatc 1130
Qy 1021 ctgagatgacagagagagcctgtgtgatatccagagaaaggaggtcccttcttctgt 1080
Db 1131 ctgagatgacagagagagcctgtgtgatatccagagaaaggaggtcccttcttctgt 1190
Qy 1081 aagctctgagagacacggtgcaatgctccctcagcaataacaccttccagctcatcatt 1140
Db 1191 aagctctgagagacacggtgcaatgctccctcagcaataacaccttccagctcatcatt 1250
Qy 1141 ggttcaaaagtgaagtgagagatttggactgtctacacgacgttgcctcagctgtgccg 1200
Db 1251 ggttcaaaagtgaagtgagagatttggactgtctacacgacgttgcctcagctgtgccg 1310
Qy 1201 ttgttcaaaagcagacacacacacacacacacacacacacacacacacacacacacacac 1260
Db 1311 ttgttcaaaagcagacacacacacacacacacacacacacacacacacacacacacacac 1370
Qy 1261 ccttctactgtggtcctcggtgttttgaacacacacacacacacacacacacacacacac 1320
Db 1371 ccttctactgtggtcctcggtgttttgaacacacacacacacacacacacacacacacac 1430
Qy 1321 aagacagctgtgacgagcggtgacatcagacagagaggtgggttgaagtgatgtatg 1380
Db 1431 aagacagctgtgacgagcggtgacatcagacagagaggtgggttgaagtgatgtatg 1490
Qy 1381 cccgtgtgataaaagaagacacacacacacacacacacacacacacacacacacacacac 1440
Db 1491 cccgtgtgataaaagaagacacacacacacacacacacacacacacacacacacacacac 1550
Qy 1441 ctggggactcctcgggagatggaagccttccgagatcttctgtcagcag 1491
Db 1551 ctggggactcctcgggagatggaagccttccgagatcttctgtcagcag 1601

```

RESULT 11

```

AAH15690
ID AAH15690 standard; cDNA; 1690 BP.
XX
AC AAH15690;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:14058.
XX
KW Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR WPI; 2001-318749/34.
XX
PR Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX
XX Claim 8; SEQ ID 14058; 2537/bp + CD ROM; English.
XX
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any special methods. AAH03166 to AAH13628 and
CC AAH1633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 1690 BP; 410 A; 469 C; 489 G; 322 T; 0 other;

```

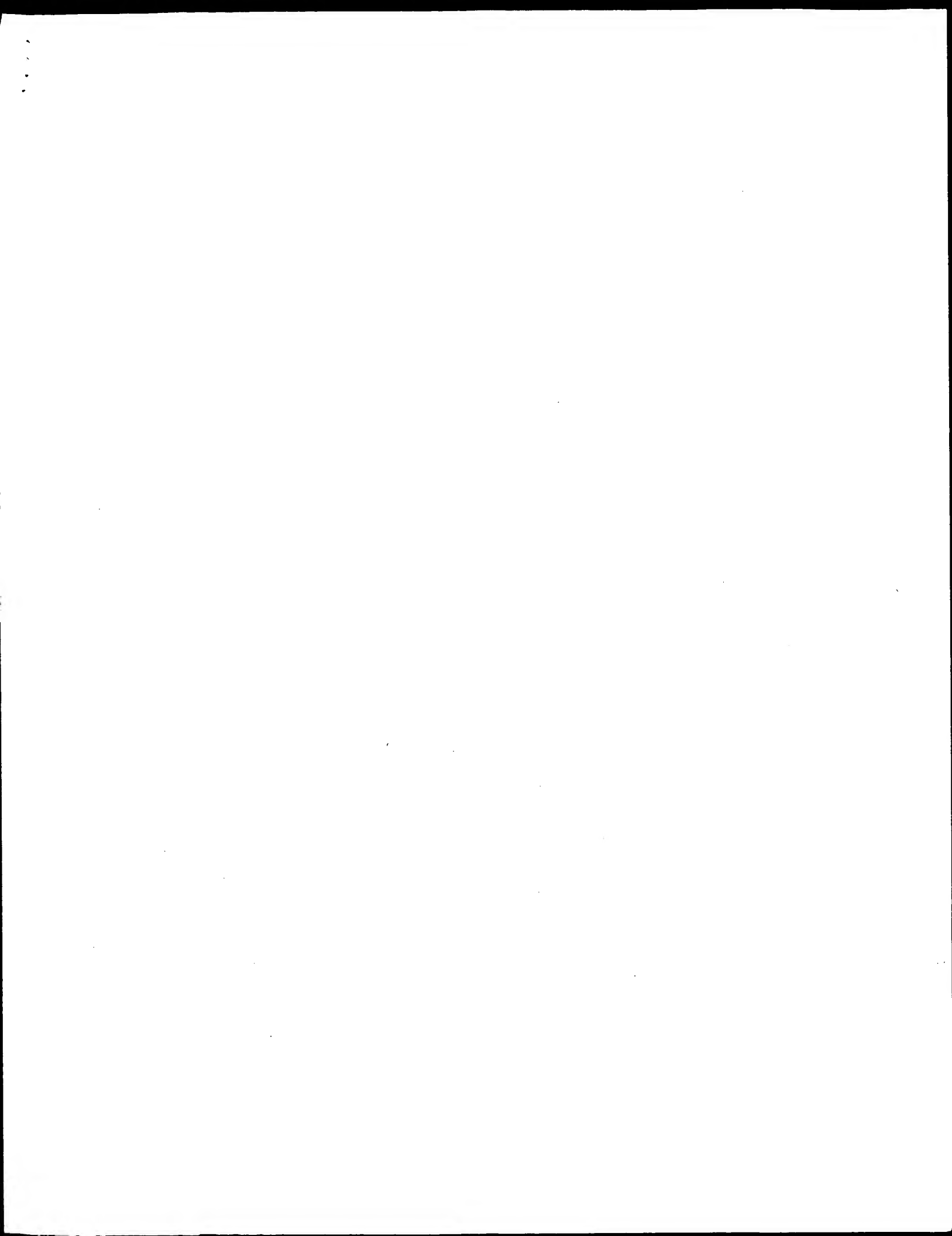
```

Query Match 74.4%; Score 1109.6; DB 22; Length 1690;
Best Local Similarity: 99.6%; Pred. No. 5,9e-293;
Matches 1112; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 atgagctgcgtcctcgggtggtggtcacccttggggcgtctgttctcgtgtcggatcc 60
Db 192 atgagctgcgtcctcgggtggtggtcacccttggggcgtctgttctcgtgtcggatcc 251

```


QY 241 acctggatgacgaactggagaagtctgtcgcagctgggccaagtacatcgtggag 300
 DB 456 acctgggagatgacgaactggagaagtcctgtcgcagctgggccaagtacatcgtggag 515
 QY 301 caaggggccacacgactgtctgtgtccatccgagcaaacctgggagcctacttggggcag 360
 DB 516 caaggggccacacgactgtctgtgtccatccgagcaaacctgggagcctacttggggcag 575
 QY 361 tatcgtctccggggttccatgtcagtcctgtatgacagagtgaaagactaacctac 420
 DB 576 tatcgtctccggggttccatgtcagtcctgtatgacagagtgaaagactaacctac 635
 QY 421 cccctaccgagcagatgagcaacccctgtgtccagagagtgctcgggagcctatgtgacg 480
 DB 636 cccctaccgagcagatgagcaacccctgtgtccagagagtgctcgggagcctatgtgacg 695
 QY 481 cactacacacagatagtttggtgcccacacacaaagaatcggtgtgtctgtgacacctgc 540
 DB 696 cactacacacagatagtttggtgcccacacacaaagaatcggtgtgtctgtgacacctgc 755
 QY 541 cggagagatgactgtctgggagagaagtttgaggagaacgcggtcacttctgtcaattat 600
 DB 756 cggagagatgactgtctgggagagaagtttgaggagaacgcggtcacttctgtcaattat 815
 QY 601 tctccaaagggaactgagttggagaagccctcaagaagaatgcccgcgcctctctgag 660
 DB 816 tctccaaagggaactgagttggagaagccctcaagaagaatgcccgcgcctctctgag 875
 QY 661 tggccaccagactatgagagcagctcagagaacaacttgttaccgagaagaacctac 720
 DB 876 tggccaccagactatgagagcagctcagagaacaacttgttaccgagaagaacctac 935
 QY 721 actc 724
 DB 936 actc 939
 RESULT 13
 AAH05058
 ID AAH05058 standard; cDNA; 840 BP.
 AC AAH05058;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA clone (5'-primer) SEQ ID NO:1893.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 OS Homo sapiens.
 XX
 EP EP1074617-A2.
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000BP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR MPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 full-length cDNAs defined in the specification, and for the detection
 and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -
 XX
 XX Claim 1: SEQ ID 1893; 2537bp + CD ROW; English.
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises at least 15 nucleotides and the combination
 CC of the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 CC
 CC Sequence 840 BP; 168 A; 246 C; 257 G; 165 T; 4 other:
 SQ
 Query Match 39.9%; Score 594.6; DB 22; Length 840;
 Best Local Similarity 96.1%; Pred. No. 2.3e-152;
 Matches 618; Conservative 0; Mismatches 23; Indels 2; Gaps 1;
 QY 1 atgagctgctgcctcgtgggtgtatcccttgggctgtcctcgtctcgtcgtac 60
 DB 192 atgagctgctgcctcgtgggtgtatcccttgggctgtcctcgtctcgtcgtac 251
 QY 61 caagctacccctcctgcccacacgtcactctcttagagagctgcctcagcaataaccagac 120
 DB 252 caagctacccctcctgcccacacgtcactctcttagagagctgcctcagcaataaccagac 311
 QY 121 aacgagctcactcccggttcgcagagcactcccccagggagaggaagatctc 180
 DB 312 aacgagctcactcccggttcgcagagcactcccccagggagaggaagatctc 371
 QY 181 atgctgcacaacaagcttcgggcccaggttcagcctcagcctcaacatggagatag 240
 DB 372 atgctgcacaacaagcttcgggcccaggttcagcctcagcctcaacatggagatag 431
 QY 241 acctgggagatgacgaactggagaagtcctgtcgcagctgggccaagtacatcgtggag 300
 DB 432 acctgggagatgacgaactggagaagtcctgtcgcagctgggccaagtacatcgtggag 491
 QY 301 caaggggccacacgactgtctgtgtccatccgagcaaacctgggagcctacttggggcag 360
 DB 492 caaggggccacacgactgtctgtgtccatccgagcaaacctgggagcctacttggggcag 551
 QY 361 tatcgtctccggggttccatgtcagtcctgtatgacagagtgaaagactaacctac 420
 DB 552 tatcgtctccggggttccatgtcagtcctgtatgacagagtgaaagactaacctac 611
 QY 421 cccctaccgagcagatgagcaacccctgtgtccagagagtgctcgggagcctatgtgacg 480
 DB 612 cccctaccgagcagatgagcaacccctgtgtccagagagtgctcgggagcctatgtgacg 671
 QY 481 cactacacacagatagtttggtgcccacacacaaagaatcggtgtgtctgtgacacctgc 540
 DB 672 cactacacacagatagtttggtgcccacacacaaagaatcggtgtgtctgtgacacctgc 731
 QY 541 cggagagatgactgtctgggagagaagtttgaggagaacgcggtcacttctgtcaattat 600
 DB 732 cggagagatgactgtctgggagagaagtttgaggagaacgcggtcacttctgtcaattat 791



Tue May 28 16:10:05 2002

us-09-667-380a-2.rspt

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2002, 14:27:40 ; Search time 122.23 Seconds
(without alignments)
703,416 Million cell updates/sec

Title: US-09-667-380A-2
Perfect score: 2759
Sequence: 1 MSCVIGVPIGLLEFLVCGS.....SESLCPNCKAFRIFAVRQ 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: SP-archaea:*
- 2: SP-bacteria:*
- 3: SP-fungi:*
- 4: SP-human:*
- 5: SP-invertebrate:*
- 6: SP-mammal:*
- 7: SP-mhc:*
- 8: SP-organelle:*
- 9: SP-phage:*
- 10: SP-plant:*
- 11: SP-rodent:*
- 12: SP-virus:*
- 13: SP-vertebrate:*
- 14: SP-unclassified:*
- 15: SP-virus:*
- 16: SP-bacteriap:*
- 17: SP-archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2756	99.9	497	4	Q9H0B8
2	2071	73.1	371	4	Q96K61
3	1987.5	72.0	434	11	Q9D2R3
4	1640.5	59.5	500	4	Q9H336
5	1613	58.5	523	13	Q985T5
6	1536.5	55.7	507	11	Q99MM6
7	968	35.1	188	11	Q920U6
8	758.5	27.5	258	13	Q985T6
9	740	26.8	258	11	Q99MM7
10	739.5	26.8	258	4	Q43692
11	646.5	23.4	253	4	Q9H3Y0
12	639	23.2	120	4	Q96T81
13	357	12.9	415	5	Q44228
14	355	12.9	266	4	Q969K2
15	349	12.6	489	11	Q9J56
16	349	12.6	489	11	Q9ET66

SEA ID NO: 2
AC, NO: Q9H0B8
Database: SP-TRMBL-19

RESULT	1	PRELIMINARY	PRT	497 AA	ALIGNMENTS
AC	Q9H0B8				
DT	01-MAR-2001 (TREMREL. 16, Created)				
DT	01-MAR-2001 (TREMREL. 16, Last sequence update)				
DT	01-DEC-2001 (TREMREL. 19, Last annotation update)				
DE	HYPOTHEtical 55.9 KDA PROTEIN.				
GN	DKFZP434B044.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
OX	NCBI_TaxID=9606;				
RP	SEQUENCE FROM N.A.				
RC	TISSUE-TESTIS;				
RA	Wambolt R., Heubner D., Mewes H.W., Gaassenhuber J., Wiemann S.;				
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL: AL136861; CAB6795.1;				
DR	HSSP: P04284; ICPE.				
DR	InterPro: IPR004043; LCCL.				
DR	InterPro: IPR001283; SCP.				
DR	Pfam: PF00188; SCP. 1.				
DR	PRINTS: PR00837; VSTPLIKE.				
DR	SMART: SM00198; SCP. 1.				
DR	PROSITE: PS01010; SCP_AGS_PRL-SC7_2; 1.				
KW	Hypothetical protein.				
SQ	SEQUENCE 497 AA; 55919 MW; 1E7ADDCAB37CA4ED CRC64;				

Query Match 99.9%; Score 2756; DB 4; Length 497;
Best Local Similarity 99.8%; Pred. No. 3.5e-259;
Matches 496; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSCVIGVPIGLLEFLVCGSGGYLLPVNTLLLEELISYQHNSHVRRAIPREDKEEL 60	
DB	1	MSCVIGVPIGLLEFLVCGSGGYLLPVNTLLLEELISYQHNSHVRRAIPREDKEEL 60	
OY	61	MLHNKLRGVQVQASNMVMTWDELKSAAMASOCITWEGPTSLTSLVIGNLGAHWR 120	

```

Db 61 MLHNKLGGOVOPASNMWYMTWDELKSAAMASOCITWEGPSTSLVSTIGONLGAHMR 120
QY 121 YRSPGFHVOSWYDEKDYTPYRSECNPMKPERCSGPMCTHTYTOIWMATTNKIGAVNMC 180
Db 121 YRSPGFHVOSWYDEKDYTPYRSECNPMKPERCSGPMCTHTYTOIWMATTNKIGAVNMC 180
QY 181 RKMTWGEWENAVYFVNCNPKGNWIGEPYKNGRPSCEPSPYSGSCNNLCYREBTY 240
Db 181 RKMTWGEWENAVYFVNCNPKGNWIGEPYKNGRPSCEPSPYSGSCNNLCYREBTY 240
QY 241 TPKEPTDENNEVEETAPRPEENHVMLOPRVMPRTPKKTSANVMTOVWCDTKMKDRCKG 300
Db 241 TPKEPTDENNEVEETAPRPEENHVMLOPRVMPRTPKKTSANVMTOVWCDTKMKDRCKG 300
QY 301 STCNRYOCPCAGCLNKAHKAIFGSLFYESSSSICRAAIHYGILDDKGLVDITRNGVPEFV 360
Db 301 STCNRYOCPCAGCLNKAHKAIFGSLFYESSSSICRAAIHYGILDDKGLVDITRNGVPEFV 360
QY 361 KSERHGVOSLSKYPSSSPMSKVKVYODDCTTYAOICPEKPRATHCPRHCPAHCKDE 420
Db 361 KSERHGVOSLSKYPSSSPMSKVKVYODDCTTYAOICPEKPRATHCPRHCPAHCKDE 420
QY 421 PSYMAVPCGNITADTSSICTAVHAGVISNESGDDVMPVDDKKTYYGSLRNGVOS 480
Db 421 PSYMAVPCGNITADTSSICTAVHAGVISNESGDDVMPVDDKKTYYGSLRNGVOS 480
QY 481 LGTPRDGKAFTFAVRQ 497
Db 481 LGTPRDGKAFTFAVRQ 497

```

```

RESULT 2
ID 096K61 PRELIMINARY; PRT: 371 AA.
AC 096K61;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE CDNA FLJ14469 FIS. CLONE MAMMA1002881. WEAKLY SIMILAR TO GLIOMA
DE PATHOGENESIS-RELATED PROTEIN.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=MAMMARY GLAND.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuno Y., Sasaki N.;
RT "NEO human cDNA sequencing project."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027395; BAB5081.1;
SQ SEQUENCE 371 AA: 42207 MW: 8023FBE14F53E85C CRC64:

```

```

Query Match
Best Local Similarity 75.1%; Score 2071; DB 4; Length 371;
Matches 369; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MSCVIGGVIPGLGLFVCGSGGILLPNVTLLELLSKYOHNSHSHVRAIPREDKEITL 60
Db 1 MSCVIGGVIPGLGLFVCGSGGILLPNVTLLELLSKYOHNSHSHVRAIPREDKEITL 60
QY 61 MLHNKLGGOVOPASNMWYMTWDELKSAAMASOCITWEGPSTSLVSTIGONLGAHMR 120
Db 61 MLHNKLGGOVOPASNMWYMTWDELKSAAMASOCITWEGPSTSLVSTIGONLGAHMR 120
QY 121 YRSPGFHVOSWYDEKDYTPYRSECNPMKPERCSGPMCTHTYTOIWMATTNKIGAVNMC 180
Db 121 YRSPGFHVOSWYDEKDYTPYRSECNPMKPERCSGPMCTHTYTOIWMATTNKIGAVNMC 180

```

```

QY 181 RKMTWGEWENAVYFVNCNPKGNWIGEPYKNGRPSCEPSPYSGSCNNLCYREBTY 240
Db 181 RKMTWGEWENAVYFVNCNPKGNWIGEPYKNGRPSCEPSPYSGSCNNLCYREBTY 240
QY 241 TPKEPTDENNEVEETAPRPEENHVMLOPRVMPRTPKKTSANVMTOVWCDTKMKDRCKG 300
Db 241 TPKEPTDENNEVEETAPRPEENHVMLOPRVMPRTPKKTSANVMTOVWCDTKMKDRCKG 300
QY 301 STCNRYOCPCAGCLNKAHKAIFGSLFYESSSSICRAAIHYGILDDKGLVDITRNGVPEFV 360
Db 301 STCNRYOCPCAGCLNKAHKAIFGSLFYESSSSICRAAIHYGILDDKGLVDITRNGVPEFV 360
QY 361 KSERHGVOSLSKYPSSSPMSKVKVYODDCTTYAOICPEKPRATHCPRHCPAHCKDE 420
Db 361 KSERHGVOSLSKYPSSSPMSKVKVYODDCTTYAOICPEKPRATHCPRHCPAHCKDE 420

```

```

RESULT 3
ID 09D2R3 PRELIMINARY; PRT: 434 AA.
AC 09D2R3;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE 1810049K24RIK PROTEIN.
GN 1810049K24RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN NCBI_TaxID=10090;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
RC MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Izawa M., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Folschmann W., Gaasterland T., Giesl C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsi G.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Guenclinc S., Hill D., Holman M., Hume D.A., Kamiya M., Lee M.H.,
RA Nordone P., Ring B., Ringwald M., Carninci P., de Bonaldo M.F.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitlaker C., Wilming L.,
RA Wyshak-Boris A., Yoshida K., Haegawa Y., Kawai H., Kohlsauki S.,
RT "functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK019034; BAB31519.1;
DR HSSP; P04284; ICEE.
DR MGD; MGI:1926142; 1810049K24RIK.
DR InterPro; IPR004043; ICCL.
DR Pfam; PF00188; SCP.1.
DR PRINTS; PR00837; V5PPLIKE.
DR SMART; SM00198; SCP.1.
DR PROSITE; PS01010; SCP_AGS_PRL_SCT_2; 1.
SQ SEQUENCE 434 AA: 48593 MW: C25067E81647AC0 CRC64:

```

```

Query Match
Best Local Similarity 72.0%; Score 1987.5; DB 11; Length 434;
Matches 348; Conservative 37; Mismatches 49; Indels 3; Gaps 1;

```

```

QY 61 MLHNKLGGOVOPASNMWYMTWDELKSAAMASOCITWEGPSTSLVSTIGONLGAHMR 120
Db 61 MLHNKLGGOVOPASNMWYMTWDELKSAAMASOCITWEGPSTSLVSTIGONLGAHMR 120

```


Db 1 MLHNLKRGVYTPPASNMENHTWDELEERSAAMAHRCMLMEHGPAGLLRSIGONLAVHNGR 60
 QY 121 YSPGPHVQSMYDEVDKDYTYPPYSECNPCWPCBRCGPMCTHTYQIWAATNTKICCAVNTC 180
 Db 61 YSPGPHVQSMYDEVDKDYTYPPYSECNPCWPCBRCGPMCTHTYQIWAATNTKICCAVNTC 120
 QY 181 RKMTWGEVMEANVYFVVCYSPKGNMIGEPAPKYNKRCSECPSPYSGSCRNILCYREETY 240
 Db 121 RNNVWGDVMEANVYFVVCYSPKGNMIGEPAPKYNKRCSECPSPYSGSCRNILCYREETY 180
 QY 241 TPKPEDEMEVETAPIPENHVLQPRVMPRTKPKTSANVMTQVYKCDTKMKDRCKG 300
 Db 181 KHPEDVMEANVESPAPRETHVWQPRVI--KTKKTPVIMFMQYVHCDTKMKDRCKG 237
 QY 301 STCNROCCPAGCLNHNKAKIFGSLFYESSSSICRAAIHYGLIDDKGLVDITRNGKVPFV 360
 Db 238 STCNROCCPAGCLNHNKAKIFGSLFYESSSSICRAAIHYGLIDDKGLVDITRNGKVPFV 297
 QY 361 KSERHGVQSLSKYKPSSEFWYKVKYVODLCYTTVAQLCPPEKATCPRTICHPACADE 420
 Db 298 KSKNGMESLSKYPSSFTVSKYETAVDCHATVAQLCPPEKATCPRTICHPACADE 357
 QY 421 PSYAPVPGINITYADTSSICKTAVHAGVYISNEGSDVDVMPVKKTYVGLRNGVQSES 480
 Db 358 PSYAPVPGINITYADTSSICKTAVHAGVYISNEGSDVDVMPVKKTYVGLRNGVQSES 417
 QY 481 LGTPPDGKAFRIFAVRQ 497
 Db 418 LNTFQNGNAFRIFAVRQ 434

RESULT 4
 09H336 PRELIMINARY; PRT; 500 AA.

AC 09H336: 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE PUTATIVE SECRETORY PROTEIN PRECURSOR (COCOACRISP).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCBL-TaxID-9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-MORTA;
 RA Zhao Y., Cao H., Jiang Y., Meng X., Zhao X., Liu D., Ding J.;
 RT "Cloning and characterization of a trypsin inhibitor-like protein gene
 in human aorta."
 RL Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-HEART, AND TESTIS;
 RA Smith D.M., Collins-Racie L.A., Lavallie E.R., Gerner L., Roberts D.J.,
 RA Marigo V.A., Copeland N.G., Jenkins N.A., McCoy J., Tabin C.J.;
 RT "A novel cysteine-rich secreted protein (CRISP) family member,
 RT Cococristp, provides insight into the process of septation in the
 RT developing chicken midbrain."
 RL Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF142573; AM643287.1;
 DR HSSP: P04284; ICFE.
 DR InterPro: IPR004043; LCCL.
 DR InterPro: IPR001283; SCP.
 DR Pfam: PF00188; SCP.1.
 DR PRINTS: PR00837; V5TPXIKR.
 DR SMART: SM00198; SCP.1.
 DR PROSITE: PS01010; SCP_AG5_PRL_SC7_2; 1.
 SEQUENCE 500 AA; 56888 MW; 203BIDC4DDA003CB CRC64;

Query Match 59.5%; Score 1640.5; DB 4; Length 500;
 Best Local Similarity 59.5%; Pred. No. 9.5e-151;

Matches 292; Conservative 74; Mismatches 112; Indels 13; Gaps 7;
 QY 13 LFLVCGSGGILLPVNTLLLELLSTKYQNE-----SHSRVRAIPREDKEELMLHNLKLR 67
 Db 14 VLFMARAIPIAMVVPVATLLEKLETKYDDEGBWIAKORGRATITNDMSQILDLHNLKLR 73
 QY 68 GQYQAPASMEYMTDDELESAAMASOCITWEGPTSLVYSGONIGAHGRTSPGFH 127
 Db 74 SQVPTASMEYMTDDELEERSAESMASCLMEHGPASLPSIGONIGAHGRTSPGFH 133
 QY 128 VQSMYDEVDKDYTYPPYSECNPCWPCBRCGPMCTHTYQIWAATNTKICCAVNTC 187
 Db 134 VQSMYDEVDKDYTYPPYSECNPCWPCBRCGPMCTHTYQIWAATNTKICCAVNTC 193
 QY 188 EWMENAYFYVCNYSKPKGNMIGEPAPKYNKRCSECPSPYSGSCRNILCYRE-ETYPPE 245
 Db 194 QIMPKAVYLVCCNYSKPKGNMIGEPAPKYNKRCSECPSPYSGSCRNILCYRE-ETYPPE 252
 QY 246 TDEMNEVEETAPIP-EEHNVVLQPRVMPRTKPKTSANVMTQVYKCDTKMKDRCKG 304
 Db 253 EEFTELEEROOSQVNDTHV--RTRSDSSRNEVISA-QQMSQIVSCVLRDQCKGTTCN 309
 QY 305 RYCCPAGCLNHNKAKIFGSLFYESSSSICRAAIHYGLIDDKGLVDITRNGKVPFV 364
 Db 310 RYCCPAGCLNHNKAKIFGSLFYESSSSICRAAIHYGLIDDKGLVDITRNGKVPFV 369
 QY 365 HGVQSLSKYKPSSEFWYKVKYVODLCYTTVAQLCPPEKATCPRTICHPACADE 424
 Db 370 NGIOTIGKYOANSFTVSKYETAVDCHATVAQLCPPEKATCPRTICHPACADE 429
 QY 425 APVFGTINITYADTSSICKTAVHAGVYISNEGSDVDVMPVKKTYVGLRNGVQSES 484
 Db 430 ARVIGTRVYSDLSICRAAHAGVYISNEGSDVDVMPVKKTYVGLRNGVQSES 488
 QY 485 RDGKAFRIFAVR 495
 Db 489 PGKAFRIFAVR 499

RESULT 5
 09H375 PRELIMINARY; PRT; 523 AA.

AC 09H375: 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE COCOACRISP.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NC NCBL-TaxID-9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Smith D.M., Collins-Racie L.A., Lavallie E.R., Gerner L., Roberts D.J.,
 RA Marigo V.A., Copeland N.G., Jenkins N.A., McCoy J., Tabin C.J.;
 RT "A novel cysteine-rich secreted protein (CRISP) family member,
 RT Cococristp, provides insight into the process of septation in the
 RT developing chicken midbrain."
 RL Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF329199; AKI6497.1;
 DR HSSP: P04284; ICFE.
 DR InterPro: IPR004043; LCCL.
 DR InterPro: IPR001283; SCP.
 DR Pfam: PF00188; SCP.1.
 DR PRINTS: PR00837; V5TPXIKR.
 DR SMART: SM00198; SCP.1.
 DR PROSITE: PS01010; SCP_AG5_PRL_SC7_2; 1.
 SEQUENCE 523 AA; 39049 MW; 3AF283C19DC3C55F CRC64;

Query Match 58.5%; Score 1613; DB 13; Length 523;
 Best Local Similarity 58.2%; Pred. No. 4.8e-148;

Matches 285; Conservative 73; Mismatches 112; Indels 20; Gaps 6;

QY 13 LFLVCGSGYLLPNTVLTLELLSKYQNHESH-----SRVRAIPREDKEEILMLHKL 66
 DB 14 LFLVAAVSAVLPNTVLTLELLSKYQNHESH-----SRVRAIPREDKEEILMLHKL 72
 QY 67 RGVOPOASNNMYMTWDELEKSAAMAASQCIWEHGPTSLVLSIGONLGAHMGRRSPGF 126
 DB 73 RGVOYPTASNNMYMTWDELEKSAAMAASQCIWEHGPTSLVLSIGONLGAHMGRRSPGF 132
 QY 127 HVQSVWDEVDYTPYPPSECPNCPGSGPMCTHYQIYMATNKGCAVNTCKRMVW 186
 DB 133 HVQAVYDEVDYTPYPPSECPNCPGSGPMCTHYQIYMATNKGCAVNTCKRMVW 192
 QY 187 GGEVWENAVYVVCNYSKGNMIGAPYKNGRPSCEPSPSGSGCRNNLCYR---EETITPK 243
 DB 193 GQIMKAVYLVVCNYSKGNMIGAPYKNGRPSCEPSPSGSGCRNNLCYR---EETITPK 252
 QY 244 PETDEMNEVE-TAPIPEENHVLQPRVMP-----TPKKTSAVNTQOVVRCDTK 294
 DB 253 EPEETNEIERQSKADATAQSPKPHSPSGSGSGSDSKNEVISTQOMQIVSCVRL 312
 QY 295 KDRCKSGTCNRYQCPAGCLNHKAKIFGSLFYESSSSICRAAIHYGLDDKGLVDITRN 354
 DB 313 RDQCKGTTCNRYQCPAGCLNHKAKIFGSLFYESSSSICRAAIHYGLDDKGLVDITRN 372
 QY 355 KVPFVYKSERHGVQSLSKYKPSSEPMVSKVYODLDCYTTVAOLCPPEKATHCPRINC 414
 DB 373 RKNFYKSYNGIQTIGKHSANSPFVSKVYVAVTCEYVQOLCPPEKATHCPRINC 432
 QY 415 AHCKDEPSVAPVFECTNTYADTSSICKTAVHAGVYSNESGGVDVMPDKKTYVSLR 474
 DB 433 HNCMOANPHYARVIGTIRNSDLSICRALHAGVYRNH-GGYVDVMPDKKTYVSLR 491
 QY 475 GVOSESLQTP 484
 DB 492 GIFSESLQTP 501

RESULT 6
 Q99MM6 PRELIMINARY; PRT; 507 AA.

AC 099MM6
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE COCACRISP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN 11
 RP SEQUENCE FROM N.A.
 RA Smith D.M., Collins-Racie L.A., Lavallie E.R., Gamber L., Roberts D.J.,
 RA Marigo V.A., Copeland N.G., Jenkins N.A., McCoy J., Tabin C.J.,
 RT "A novel cysteine-rich secreted protein (CRISP) family member,
 RT Cocacrisp, provides insight into the process of septation in the
 RT developing chicken midbrain."
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF229198; AAK16496.1;
 DR HSSP: P04284; ICPE.
 DR InterPro: IPR004043; LCCL.
 DR InterPro: IPR001283; SCP.
 DR Pfam: PF00188; SCP.1.
 DR PRINTS: PR00837; V5TPXLIKE.
 DR SMART: SM00198; SCP.1.
 DR PROSITE: PS01010; SCP_AGS_PRL_SC7_2; 1.
 SO SEQUENCE 507 AA; 57882 MW; 554C8DFE29A50A02 CRC64;

Query Match 55.7%; Score 1536.5; DB 11; Length 507;
 Best Local Similarity 55.4%; Pred. No. 1.2e-140;
 Matches 274; Conservative 74; Mismatches 112; Indels 35; Gaps 7;

QY 13 LFLVCGSGYLLPNTVLTLELLSKYQNHESH-----SRVRAIPREDKEEILMLHKL 66
 DB 14 LFLVAAVSAVLPNTVLTLELLSKYQNHESH-----SRVRAIPREDKEEILMLHKL 72
 QY 66 LGVOPOASNNMYMTWDELEKSAAMAASQCIWEHGPTSLVLSIGONLGAHMGRRSPGF 125
 DB 72 LGVOYPTASNNMYMTWDELEKSAAMAASQCIWEHGPTSLVLSIGONLGAHMGRRSPGF 131
 QY 126 HVQSVWDEVDYTPYPPSECPNCPGSGPMCTHYQIYMATNKGCAVNTCKRMVW 185
 DB 132 HVQAVYDEVDYTPYPPSECPNCPGSGPMCTHYQIYMATNKGCAVNTCKRMVW 191
 QY 186 GGEVWENAVYVVCNYSKGNMIGAPYKNGRPSCEPSPSGSGCRNNLCYR---EETITPK 242
 DB 192 GQIMKAVYLVVCNYSKGNMIGAPYKNGRPSCEPSPSGSGCRNNLCYR---EETITPK 251
 QY 243 PETDEMNEVE-TAPIPEENHVLQPRVMP-----TPKKTSAVNTQOVVRCDTK 293
 DB 252 EPEETNEIERQSKADATAQSPKPHSPSGSGSGSDSKNEVISTQOMQIVSCVRL 312
 QY 294 KDRCKSGTCNRYQCPAGCLNHKAKIFGSLFYESSSSICRAAIHYGLDDKGLVDITRN 353
 DB 299 LRDQCKGTTCNRYQCPAGCLNHKAKIFGSLFYESSSSICRAAIHYGLDDKGLVDITRN 358
 QY 354 KVPFVYKSERHGVQSLSKYKPSSEPMVSKVYODLDCYTTVAOLCPPEKATHCPRINC 413
 DB 359 GRKHFYKSYNGIQTIGKHSANSPFVSKVYVAVTCEYVQOLCPPEKATHCPRINC 418
 QY 414 AHCKDEPSVAPVFECTNTYADTSSICKTAVHAGVYSNESGGVDVMPDKKTYVSLR 473
 DB 419 HNCMOANPHYARVIGTIRNSDLSICRALHAGVYRNH-GGYVDVMPDKKTYVSLR 492
 QY 474 GVOSESLQTP 488
 DB 478 NGIFSESLQTP 492

RESULT 7
 Q92006 PRELIMINARY; PRT; 188 AA.

AC 092006
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE LATE GESTATION LUNG PROTEIN 1.
 GN LGI1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxID=10116;
 RN 11
 RP SEQUENCE FROM N.A.
 RA STRAIN-WISTAR; TISSUE-LUNG;
 RX MEDLINE-99292450; PubMed-10362728;
 RA Kaplan F., Ledoux P., Kaasmali F.O., Gagnon S., Post M., Koehler D.,
 RA Delmont J., Sweezey N.B.;
 RT "A novel developmentally regulated gene in lung mesenchyme: homology
 RT to a tumor derived trypsin inhibitor."
 RL Am. J. Physiol. 276:L1027-L1036(1999).
 DR EMBL: AF10674; AAD16986.1;
 DR HSSP: P04284; ICPE.
 DR InterPro: IPR001283; SCP.
 DR Pfam: PF00188; SCP.1.
 DR PRINTS: PR00837; V5TPXLIKE.
 DR SMART: SM00198; SCP.1.
 DR PROSITE: PS01010; SCP_AGS_PRL_SC7_2; 1.
 SO SEQUENCE 188 AA; 21907 MW; 5B6C7715360BCE CRC64;

Query Match 35.1%; Score 968; DB 11; Length 188;
 Best Local Similarity 86.5%; Pred. No. 5.9e-86;
 Matches 160; Conservative 10; Mismatches 15; Indels 0; Gaps 0;

QY 61 MLHNKRGVOVQASNMETWDELEKSAAMASOCITWEHPTSLIVSICGNLGAHWGR 120
 |||||
 Db 1 MLHNKRGVOVPPASNMETWDELEKSAAMASOCITWEHPTSLIVSICGNLGAHWGR 60
 QY 121 YRSPGFHVQSWYDEVKDYTPYPSECPNCPGRCSPGCTHTOTVWATNTKIGCAVNTC 180
 |||||
 Db 61 YRSPGFHVQSWYDEVKDYTPYPSECPNCPGRCSPGCTHTOTVWATNTKIGCAVNTC 120
 QY 181 RKMTVGEVWENAVYFVCNYSRPGNMWIGEPYKNGRPGCEPSPYSGSCNNLCYREETY 240
 |||||
 Db 121 RSMYVGDIMENAVYLVVCNYSRPGNMWIGEPYKNGRPGCEPSPYSGSCNNLCYREETY 180
 QY 241 TPKE 245
 |||||
 Db 181 HOKPE 185

RESULT 8
 098ST6 PRELIMINARY: PRT: 258 AA.
 AC 098ST6:
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE SUGARCRISP.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauiria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_Taxid=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-STAGE 26 LIMB BUD;
 RX MEDLINE=21184121; PubMed=11287197;
 RA Smith D.M., Collins-Racie L.A., Marigo V.A., Roberts D.J., Davis N.M.,
 RA Hartmann C., Schweitzer R., Lavallie E.R., Gamer L., McCoy J.,
 RA Tabin C.J.;
 RT "Cloning and expression of a novel cysteine-rich secreted protein
 RT family member expressed in thyroid and pancreatic mesoderm within the
 RT chicken embryo."
 RL Mech. Dev. 102:223-226(2001).
 DR EMBL: AF329195; AAK16493.1; -.
 DR HSSP: P04284; ICPE.
 DR InterPro: IPR001283; SCP.
 DR Pfam: PF00188; SCP, 1.
 DR PRINTS: PRO0837; V5TPXLIKE.
 DR SMART: SM00198; SCP, 1.
 DR PROSITE: PS01010; SCP_AG5_PRL_SC7_2; 1.
 DR PROSITE: PS01010; SCP_AG5_PRL_SC7_2; 1.
 SQ SEQUENCE 258 AA; 29238 MW; 3FD85945B92EF35 CRC64;

Query Match 27.5%; Score 758.5; DB 13; Length 258;
 Best Local Similarity 56.2%; Pred. No. 2,1e-65;
 Matches 140; Conservative 29; Mismatches 57; Indels 23; Gaps 4;

QY 1 MSCVGLGVIPGLFLVNGSOCIILPNTLLLELSKIQ-----HNSR-----HSR 46
 |||||
 Db 7 ISCVF-----LRSILCETSAIVLPNSTDL--LLSNNTFIDETALAAHDSAKIPKAR 57
 QY 47 VRAIPREDKEEILMLHNKRGVOVQASNMETWDELEKSAAMASOCITWEHPTSL 106
 |||||
 Db 58 RKRYSQNDMAIILDYHNOVGKVPAPASNMETWDELEKSAAMASOCITWEHPTSL 117
 QY 107 LVSIQGNLGAHWGRYSPGFHVQSWYDEVKDYTPYPSECPNCPGRCSPGCTHTOTV 166
 |||||
 Db 118 LRFIQONTSVYTGKRYRSLIQLVKFWYDEVKDYTPYPSECPNCPGRCSPGCTHTOTV 177
 QY 167 WATNTKIGCAVNTCGRKMTVGEVWENAVYFVCNYSRPGNMWIGEPYKNGRPGCEPSPY 226
 |||||
 Db 178 WATNTKIGCAVNTCGRKMTVGEVWENAVYFVCNYSRPGNMWIGEPYKNGRPGCEPSPY 237
 QY 227 GSCNNLCY 235

Db 238 GSCNNLCY 246
 |||||

RESULT 9
 099M7 PRELIMINARY: PRT: 258 AA.
 AC 099M7:
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE SUGARCRISP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21184121; PubMed=11287197;
 RA Smith D.M., Collins-Racie L.A., Marigo V.A., Roberts D.J., Davis N.M.,
 RA Hartmann C., Schweitzer R., Lavallie E.R., Gamer L., McCoy J.,
 RA Tabin C.J.;
 RT "Cloning and expression of a novel cysteine-rich secreted protein
 RT family member expressed in thyroid and pancreatic mesoderm within the
 RT chicken embryo."
 RL Mech. Dev. 102:223-226(2001).
 DR EMBL: AF329195; AAK16494.1; -.
 DR HSSP: P04284; ICPE.
 DR InterPro: IPR001283; SCP.
 DR Pfam: PF00188; SCP, 1.
 DR PRINTS: PRO0837; V5TPXLIKE.
 DR SMART: SM00198; SCP, 1.
 DR PROSITE: PS01010; SCP_AG5_PRL_SC7_2; 1.
 DR PROSITE: PS01010; SCP_AG5_PRL_SC7_2; 1.
 SQ SEQUENCE 258 AA; 29150 MW; 2FA9949A2562640 CRC64;

Query Match 26.8%; Score 740; DB 11; Length 258;
 Best Local Similarity 55.7%; Pred. No. 1.3e-63;
 Matches 131; Conservative 31; Mismatches 61; Indels 12; Gaps 1;

QY 13 LFLVNGSOCIILPNTLLLELSKYQHNS-----HSRRRAIPREDKEIL 60
 |||||
 Db 12 LLSLCEHNTVLLNPYDSSLPANNFTDTEPALSTPLESADIPKARRKYSQNDMAIIL 71
 QY 61 MLHNKRGVOVQASNMETWDELEKSAAMASOCITWEHPTSLIVSICGNLGAHWGR 120
 |||||
 Db 72 DYHNOVGKVPAPASNMETWDELEKSAAMASOCITWEHPTSLIVSICGNLGAHWGR 131
 QY 121 YRSPGFHVQSWYDEVKDYTPYPSECPNCPGRCSPGCTHTOTVWATNTKIGCAVNTC 180
 |||||
 Db 132 YRSILOLVKFWYDEVKDYTPYPSECPNCPGRCSPGCTHTOTVWATNTKIGCAVNTC 191
 QY 181 RKMTVGEVWENAVYFVCNYSRPGNMWIGEPYKNGRPGCEPSPYSGSCNNLCY 235
 |||||
 Db 192 QNMVWGSVVRRAVYLVVCNYSRPGNMWIGEPYKNGRPGCEPSPYSGSCNNLCY 246
 RESULT 10
 043692 PRELIMINARY: PRT: 258 AA.
 AC 043692:
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE 25 KDA TRYPSIN INHIBITOR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=98146272; PubMed=9473672;
 RX Yamakawa T., Miyata S., Ogawa N., Koshikawa N., Yasumitsu H.,

RA Kanamori T., Miyazaki K.;
 RT "cDNA cloning of a novel trypsin inhibitor with similarity to
 RT pathogenesis-related proteins, and its frequent expression in human
 RT brain cancer cells.";
 RL Blochim. Biophys. Acta 1395:202-208(1998).
 DR EMBL; D45027; BAA25066.1; -.
 DR HSSP; P04284; ICPE.
 DR InterPro: IPR001283; SCP.
 DR Pfam: PF00188; SCP; 1.
 DR PRINTS; PR00837; V5TPXLIKE.
 DR SMART; SM00198; SCP; 1.
 DR PROSITE; PS01010; SCP_AGS_P1_SCF_2; 1.
 SQ SEQUENCE 258 AA; 29065 MW; 1915A5831637795F CRC64;

Query Match 26.8%; Score 739.5; DB 4; Length 258;
 Best Local Similarity 56.4%; Pred. No. 1.5e-63;
 Matches 133; Conservative 30; Mismatches 60; Indels 13; Gaps 2;

QY 13 LIF-LVCGSGYLLPVTLLLEELSKYQHNES-----HSRYRAIPREDKEEI 59
 DB 11 LIFSLICEASTYVLLNSTDSSPTNPFIDIEALKAQLDSADIPAKRRKRTISQNDMAI 70
 QY 60 LMLHNLKRGQVOPQASNMEXMTWDELEKSAAMASQCIWEHPTSLVSTIGNLGAHWG 119
 DB 71 LDYHNVGRKGKPPPAANMEYMWDEMLAKSADAMATCTIMDGPSTLLRFLGQNLSTVETG 130
 QY 120 KRSPGFHVSQWYDEKQDTVPYSECNWCPERGSCPMCTHYQIVATNTKKICAVNT 179
 DB 131 RRRSILQIVKPMWDEKDAFPYPODCNPRCPRCFGPCTHYTQMWATSNRICALHT 190
 QY 180 CRMTWGEVWENAVFYVCNYSKPGNMIGEAIFYKNGRPGCECPSPSYGSGCRNNLCY 235
 DB 191 CQMMNVGSMVWIRAVYLVGNVAPKGMWIGEAIFYKNGRPGCECPSPSYGSGCTNLCF 246

RESULT 11
 Q9H3Y0 PRELIMINARY; PRT; 253 AA.
 AC Q9H3Y0;
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE DJ881L22.3 (NOVEL PROTEIN SIMILAR TO A TRYPSIN INHIBITOR).
 GN DJ881L22.3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lloyd D.;
 RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AL117382; CAC04190.1; -.
 DR HSSP; P04284; ICPE.
 DR InterPro: IPR001283; SCP.
 DR Pfam; PF00188; SCP; 1.
 DR PRINTS; PR00837; V5TPXLIKE.
 DR SMART; SM00198; SCP; 1.
 SQ SEQUENCE 253 AA; 28605 MW; 28976C6F32E3867E CRC64;

Query Match 23.4%; Score 646.5; DB 4; Length 253;
 Best Local Similarity 47.7%; Pred. No. 1.6e-54;
 Matches 114; Conservative 38; Mismatches 66; Indels 21; Gaps 4;
 QY 12 GLLEFLVCGS-QGYLLPNT-----LIEEL-LSKYQHNESHVRRAIPREDKE 57
 DB 12 GLLEFLVCGS-QGYLLPNT-----LIEEL-LSKYQHNESHVRRAIPREDKE 57
 QY 58 EILMLANKRGQVOPQASNMEXMTWDELEKSAAMASQCIWEHPTSLVSTIGNLGAHWG 117
 DB 65 ALDYNHNRASVYPPAANMEYMWDRRLAARAAEWATQCIWAHGPSQIMRYGQNLSTH 124

QY 118 MGRYSPGFHVSQWYDEKQDTVPYSECNWCPERGSCPMCTHYQIVATNTKKICAV 177
 DB 125 SGQYRSYVDMKMSSEKWHLPAPDCNPHCPMKCDGFTCSHYQMWASSNRIGCAI 184
 QY 178 NTCRMTWGEVWENAVFYVCNYSKPGNMIGEAIFYKNGRPGCECPSPSYGSGCRNNLCY 236
 DB 185 HTCSISVWGTWIRAVYLVGNVAPKGMWIGEAIFYKNGRPGCECPSPSYGSGCRNNLCY 243

RESULT 12
 Q96IB1 PRELIMINARY; PRT; 120 AA.
 AC Q96IB1;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DE HYPOTHETICAL 13.5 KDA PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tissue-MELANOMA;
 RA Strausberg R.;
 RL Submitted (May-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; BC007689; AA07689.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 120 AA; 13497 MW; D0B40A3149B1099C CRC64;

Query Match 23.2%; Score 639; DB 4; Length 120;
 Best Local Similarity 100.0%; Pred. No. 3.1e-54;
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSCVLGGYIPGLLEFLVCGSGYLLPVTLLLEELSKYQHNESHVRRAIPREDKEEI 60
 DB 1 MSCVLGGYIPGLLEFLVCGSGYLLPVTLLLEELSKYQHNESHVRRAIPREDKEEI 60
 QY 61 MLHNLKRGQVOPQASNMEXMTWDELEKSAAMASQCIWEHPTSLVSTIGNLGAHWGR 120
 DB 61 MLHNLKRGQVOPQASNMEXMTWDELEKSAAMASQCIWEHPTSLVSTIGNLGAHWGR 120

RESULT 13
 O44228 PRELIMINARY; PRT; 415 AA.
 AC O44228;
 DT 01-JUN-1998 (TREMblrel. 06, Created)
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
 DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE HRTT-1 PROTEIN.
 GN HRTT-1.
 OS Halocynthia roretzi (Sea squirt).
 OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
 OC Stolidobranchia; Pyridae; Halocynthia.
 OX NCBI_TaxID=7729;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-98267369; PubMed-9601991;
 RA Holta K., Takahashi H., Satoh N.;
 RT "Expression of an ascidian gene in the tip of the tail of tail-bud-
 RT stage embryos.";
 RL Dev. Genes Evol. 208:164-167(1998).
 DR EMBL; AB009609; BAA24011.1; -.
 DR HSSP; P04284; ICPE.
 DR InterPro: IPR001283; SCP.
 DR InterPro: IPR000884; TSP1.
 DR Pfam; PF00188; SCP; 1.
 DR Pfam; PF00090; TSP_1; 1.
 DR PRINTS; PR00837; V5TPXLIKE.
 DR SMART; SM00198; SCP; 1.

DR SMART: SM00209: TSPL 1
DR PROSITE: PS01009: SCP_A65_PRL_SC7_1; 1.
DR PROSITE: PS01010: SCP_A65_PRL_SC7_2; 1.
DR PROSITE: PS50092: TSPL 1.
SQ SEQUENCE 415 AA; 47365 MW; 2266B3F19DEDD3 CRC64;

Query Match
Best Local Similarity 12.9%; Score 357; DB 5; Length 415;
Matches 110; Conservative 58; Mismatches 128; Indels 96; Gaps 18;

QY 24 LIPNTLLEELLSKY-----QHNSHSRVRA-----IREDEKEE-----LM 61
DB 74 ILSEKVEEITARDIGLDPANETFS-VKAAVAKIIEGKIMDKIVLKPERMALLR 132
QY 62 LHNKRGVQPOASMEYTWDELEKSAAMASQIWEHPT---SLVSIIGNLGAHW 118
DB 133 THNARRAAOPKAAKMRMTWDELEKLAVALYSRKCIEHNPRIKHSFEYVGENLFIST 192
QY 119 GRYRSGPF---HYOSVDEVDYTYPPYSECNPMCPERCS-GPMCTHYQIWMATTKNG 174
DB 193 GYALFPLSMKHAIVEAMDEKQYDE-----TKKQKQKMGCHYTVWADTFEMG 243
QY 175 CAVNTCRKMTYGEWENAVYFCNYSPEKGMIGEAPYKNGRPGSECPYSGSCRNLC 234
DB 244 CGYTRCSIDYRGRKMKAILLVNCGGNY-PTHPEVTAPSCSKAPT--DICRRNLC 300
QY 235 YREETYPKPTDEMEY--ETADIPBENHYVLOPRVWRPTPKTSAVNYMTQVRCDF 292
DB 301 -----NNVIRDLRLDKRDKIMSEMTW--SSCSKSGVSGSTRERCNT 343
QY 293 KMKDRCKG---STCNRYOCPAGCLNKKKIRFSLFEYSSSICRAIHYIGLIDKGL 347
DB 344 FVPEDCKDFSEVFCYKRC-----KAAKGN-----GGSFYNIWMNG-- 384
QY 348 VDIRNKPVEPFYKSEHGVOSISKYPSSEF 379
DB 385 -DKLLKSLQ-----QALQKHLGFSF 405

RESULT 14
Q969K2 PRELIMINARY; PRT; 266 AA.
AC Q969K2; 01-DEC-2001 (Tremblrel, 19, Created)
DT 01-DEC-2001 (Tremblrel, 19, last sequence update)
DE GLIOMA PATHOGENESIS-RELATED PROTEIN (UNKNOWN) (PROTEIN FOR
GN RTVP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wu J., Zhang B., Zhou Y., Peng X., Yuan J., Qiang B.;
RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
RC TISSUE=RENAL ADENOCARCINOMA;
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF400440; AAK92489.1;
DR EMBL; BC012510; AAH12510.1;
SQ SEQUENCE 266 AA; 30366 MW; C0466C085EFB2B7A CRC64;

Query Match
Best Local Similarity 12.9%; Score 355; DB 4; Length 266;
Matches 82; Conservative 33; Mismatches 75; Indels 40; Gaps 12;
QY 25 LPNVTLLLEELLSKYOHNSHSRVRAIPRED-KEIILNHNKRGVQPOASMEYMTWD 83

DB 5 LATIAMWSEFVSNYSHT---ANILPDIEENEDFIDCVRIHNNKFESEVKPTASDMLYMTWD 61
QY 84 DELEKSAAMASQIWEHPT---PTSL---LVSIIGNLGAHWCRVYSPGHQV-----W 131
DB 62 PALAQIAKAMASNGQFSHNTRPKPKHLHFNFTSLGENT--W-TGSVPLEFVSATNW 117
QY 132 YDEKVDYTYPPYSCNPMCPERCSGPMCTHYQIWMATTKNGICAVNTCRKMTYGEW 191
DB 118 YDEIYDQYDF-----KTRICKKVCCHYTVWADSYKVCAGVQFCPKYSGF-DALS 166
QY 192 NAVYFCNYSPEKGN---WIGAPYKNGRPGSECPYSGSCRNLCYREE 238
DB 167 NGAHFICNYPGGNYPTW---PYKRGATCSACPNN--DKCLDNLGVNRQ 210

RESULT 15
Q9J56 PRELIMINARY; PRT; 489 AA.
AC Q9J56;
DT 01-OCT-2000 (Tremblrel, 15, Created)
DT 01-OCT-2000 (Tremblrel, 15, last sequence update)
DE CYSTEINE-RICH PROTEASE INHIBITOR.
GN CRIP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10900;
RN [1]
RP SEQUENCE FROM N.A.
RA Jang J.S., Hahn Y., Chung J.H.;
RT Identification of novel mouse cysteine-rich protease inhibitor
gene."
RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB046537; BAB03398.1; -.
DR HSSP; P04284; ICEE.
DR InterPro; IPR001283; SCP.
DR Pfam; PF00188; SCP; 1.
DR PRINTS; PR00837; V5TPXLIKE.
DR SMART; SM00198; SCP; 1.
DR PROSITE; PS01009; SCP_A65_PRL_SC7_1; 1.
DR PROSITE; PS01010; SCP_A65_PRL_SC7_2; 1.
KW Protease.
SQ SEQUENCE 489 AA; 52676 MW; 01C207BE12E3CB9D CRC64;

Query Match
Best Local Similarity 12.6%; Score 349; DB 11; Length 489;
Matches 77; Conservative 28; Mismatches 86; Indels 16; Gaps 6;
QY 50 ALPRDKKEIILNHNKLGVOPOASMEYTWDELEKSAAMASQIWEHPTSLVLS 109
DB 20 ALTEDEKOTMVDLHNOYTAOVSPPASDMLQKRWDELAFAKAAOKVWGNHER--GR 77
QY 110 IGNLGAHWGRYRSGPFHYOSVDEVDYTYPPYSECNPMCPERCSGPMCTHYQIWMAT 169
DB 78 RGENLFAITDEGMVPLAVGAWMHEHEVYFNFS-TATCDP-----NOMCHYTVWWSK 129
QY 170 TNKIGCAVNTCRKMTYGEWENAVYFCNYSPEKGMIGEAPYKNGRPGSECPYSGSC 229
DB 130 TERIGCGHFE--TLQGEVANILHLVCNYPGNVYKGRKPYOEGTPSCGCPFGY--SC 165
QY 230 RNNLCYREETYPKPTDEMEYVETAP 256
DB 186 ENSLC--EPMRNPBKADSPRVTREVP 210

Search completed: May 25, 2002, 14:38:41
Job time: 661 sec

copy b.1 + 2

GenCore version 4.5
 Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2002, 10:30:04 ; Search time 2776.47 Seconds

(without alignments)
 11337.819 Million cell updates/sec

Title: US-09-667-380A-1
 Perfect score: 1491
 Sequence: 1 atgagctgcgtctcctgggtgg.....ggatcttgcgtcagcagcag 1491

Scoring table: IDENTITY_NUC
 Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
 Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : GenEmbl:*
 1: gb.ba:*
 2: gb.htg:*
 3: gb.ln:*
 4: gb.om:*
 5: gb.ov:*
 6: gb.pat:*
 7: gb.ph:*
 8: gb.pl:*
 9: gb.pr:*
 10: gb.ro:*
 11: gb.sts:*
 12: gb.sy:*
 13: gb.un:*
 14: gb.vi:*
 15: em.ba:*
 16: em.fun:*
 17: em.hum:*
 18: em.ln:*
 19: em.mn:*
 20: em.om:*
 21: em.or:*
 22: em.ov:*
 23: em.pat:*
 24: em.ph:*
 25: em.pl:*
 26: em.ro:*
 27: em.sts:*
 28: em.un:*
 29: em.vi:*
 30: em.htg.hum:*
 31: em.htg.ln:*
 32: em.htg.other:*
 33: em.htgo.ln:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query
 No. Score Match length DB ID Description

1	1491	100.0	1491	6	AX101173
2	1491	100.0	2272	6	AX101175
3	1487.8	99.8	4574	6	AX086850
4	1487.8	99.8	4574	6	AX285067
5	1487.8	99.8	4877	6	AX285067
6	1487.8	99.8	4877	6	AX285068
7	1487.8	99.8	4877	6	AX285079
8	1486.2	99.7	2400	6	AX235369
9	1474.8	98.9	2403	6	AX235373
10	1452.6	97.4	2412	6	AX235375
11	1393.8	93.5	1669	6	AX235371
12	1286.2	86.3	2305	6	AX235367
13	1109.6	74.4	1690	9	AX027395
14	974.2	65.3	3052	10	AF109674
15	484.4	32.5	2664	6	AX250094
16	484.4	32.5	2664	9	AF142573
17	484.4	32.5	2667	9	AF329197
18	484.4	32.5	2962	9	BC020514
19	458.6	30.8	2543	10	AF329198
20	441.8	29.6	2382	5	AF329199
21	362	24.3	857	9	BC007689
22	269.4	18.1	1498	6	AX250025
23	240	16.1	174714	9	AC025280
24	238.4	15.0	1365	10	AF329196
25	238.4	15.7	1394	5	AF329195
26	218.6	14.7	594	23	E09937
27	218.6	14.7	1440	9	D45027
28	146	9.8	181161	2	AC092347
29	146	9.8	181418	2	AC092341
30	126.4	8.5	128	6	AX318722
31	89.2	6.0	67263	2	AC101056
32	79.8	5.4	1877	6	AX136261
33	79.8	5.4	2026	9	BC022399
34	75.4	5.1	720	6	AX136546
35	74.6	5.0	970	6	AR026670
36	74.6	5.0	970	6	AR097602
37	73	4.9	184558	2	AC020570
38	73	4.9	185123	9	AC100782
39	69.8	4.7	122	10	AY033439
40	68.8	4.6	181267	2	AL627168
41	68.6	4.6	167819	2	AC103528
42	65.2	4.4	160128	2	AC093982
43	64.2	4.3	2415	10	AB046537
44	62.8	4.2	156235	9	AC084336
45	62.8	4.2	175743	9	AC011632

ALIGNMENTS

RESULT 1
 AX101173
 LOCUS AX101173 1491 bp
 DEFINITION Sequence 1 from Patent WO0121651.
 ACCESSION AX101173
 VERSION AX101173.1 GI:13620006
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 1 (bases 1 to 1491)
 Donoho, G., Turner, C.A., Waller, F., Nehls, M., Friedrich, G.,
 Zambrowicz, B. and Sands, A.T.
 Novel human protease inhibitor-like proteins and polynucleotides
 encoding the same
 Patent: WO 0121651-A 1 29-MAR-2001;
 JOURNAL Lexicon Genetics Incorporated (US)
 LOCATION/Qualifiers
 1..1491
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

FEATURES

source

BASE COUNT 368 a 409 c 426 g 288 t

QY 121 aacgagctcactcccggttcgagagccatccccagggagagagagagatctc 180
 DB 336 AACGAGCTCCTCCCGGCTCCGAGAGCCATCCCAAGGAGAGAGATCTCC 395
 QY 181 atgctgcaacaagcttcggggccaggtgtcagctcagctccacaatgagatc 240
 DB 396 ATGCTGCAACAAGCTTCGGGGCCAGGTGTCAAGCTCCACAATGAGATCA 455
 QY 241 acctggatgacgaactggagaagtcgtcagcgttggccagtcagtgtcagg 300
 DB 456 ACCTGGATGACGAAGTCTGCTGACGCTGGGCCAGTCAATCTGGGAG 515
 QY 301 caagggcccaacagctcgtgtgtccatccggcagaaccttggcgctacacg 360
 DB 516 CACGGGCCACACAGTCTGTGTCTCATCGGCGACAACTGGGCGCTACTGG 575
 QY 361 tctgctctccgggttcacatgtcagtcgtctgtatgacgaggtgaaagac 420
 DB 576 TATCGCTCCGGGTTCCATGTGAGTCTGTGATGACGAGGTGAAGACTAC 635
 QY 421 ccctaccgagagtgcaacccctgtgtccagagaggtgtctcgggcttatgt 480
 DB 636 CCTTACCCGAGCGAGTCAACCCCTGTGTCTGACAGAGTGTCTGGGCTAT 695
 QY 481 cactacaacagaatgttggggccaccacaacagaatcgtgtgtgtgaaacac 540
 DB 696 CACTACACACAGATAGTTGGCCACCACCAACAGATCGTTGTCTGTGAAC 755
 QY 541 cggagatgactgtctgggagagatttgggagagcggtctacttgtcgaatt 600
 DB 756 CGGAGATGACTGCTCTGGGAGAAATTGGAGAACCGGCTACTTGTCTCAAT 815
 QY 601 tctccaaagggagacgtgtgtgtgagagagcccttacaagaatggcgccct 660
 DB 816 TCTCCAAAGGGGAGATGATGATGAGAGCCCTTACAGATGCGGCGCTCTG 875
 QY 661 tggccaccagctatgtgagagcagctgtcaggaacaacttgttaccgagaag 720
 DB 876 TGCCACCCAGCTATGAGAGCGCAGTGCAGAACACTGTGTACCGAAGAACT 935
 QY 721 actccaaacactgaacagagagatgaatgaatgtgaaacggtctccatct 780
 DB 936 ACTCCAAACCTGAACGAGAGATGAATGATGAGTGAAGAGGCTCCATCTC 995
 QY 781 aacatgttgggtccaaacgaggtgtatgagaccacaagcccaagaacact 840
 DB 996 AACATGTGTGGCTCCAAACGAGGTGATGAGACCCCAAGCCCAAGAAACT 1055
 QY 841 gtcacatacatgacccaagtcgtcagaatgtgtacacccaagaatgaagag 900
 DB 1056 GTCAACTATCATGACCCAAAGTGTGATGATGACACCAAGATGAAGAGG 1115
 QY 901 tccacgtgtacaggtaccagtcgcccgagcggtcgtctgaaacaaaggcgaa 960
 DB 1116 TCCACGTGTACAGGTACAGTGCACAGAGGCTGCTGAACCAACAGGCAAT 1175
 QY 961 ggaagctgtcttatagaagctgtctacagcatgtccgagcgccatccact 1020
 DB 1176 GGAAGTCTTCTATGAAGAGTCTGCTATGATGATGCGCGCGCATACAG 1235
 QY 1021 ctgtatgacaaaggagcgtgtgtgatacaacaggaaggaaggtccctct 1080
 DB 1236 CTGTATGACAAAGGAGCGCTGTGTGATATCACACAGGAAGGAGTCCCT 1295
 QY 1081 aagctgtgagagagagcggtcagtcctcctcagcaatacaacacttccag 1140
 DB 1296 AAGCTGTGAGAGACAGCGGTGAGTCTCTCAGCAAAATCAAACTTCAAG 1355
 QY 1141 gtgtcaaaagtgaaggtcagagatttggactgttaccagacggttgcact 1200
 DB 1356 GTGTCAAAAGTGAAGGTGAGATTTGGACGTCTACAGACGCTGTGCTG 1415
 QY 1201 ttgtgaagcagcaactcactgtcccaagaatcctgttccggacactgtca 1260

DB 1416 TTTGAAAGCCAGCACTACCTGCCCCAAGATCATTTGCTGGGACACTCAAG 1475
 QY 1261 ccttcactactgttcggtgttttggaaaccaactcatgtcaggtacttaacat 1320
 DB 1476 CCTTCTACTTGGGTCCGGTGTGGAAACCAACTATGATGATACCTTAAGCAT 1335
 QY 1321 aagcaagctgtgacagcgaggtatcatcagcaacagagtgaggtgtgacgt 1380
 DB 1536 AAGACAGCTGTGACAGCGGAGTCAATGACGAAGAGAGTGGGGTGAAGT 1595
 QY 1381 cccgtgtgataaagaagaacacacgtgtgtgtcgtcaggaatgagttcgtga 1440
 DB 1596 CCGTGTATTAAGAAAGAACCTACAGTGGGCTCGCTCAGGAATGAGTCAAT 1655
 QY 1441 ctggagactcctcgtgagatgagaaagccttcggatcttctgtcagga 1491
 DB 1656 CTGGGACTCTCGGATGAGAAAGGCTTCCGATCTTGTCTGACAGCAG 1706

RESULT 3

LOCUS AX086850 4574 bp DNA linear PAT 09-MAR-2001
 DEFINITION Sequence 802 from Patent WO0112659.
 ACCESSION AX086850
 VERSION AX086850.1 GI:13276048
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

1 (bases 1 to 4574)
 Wilmann, S.
 Human dna sequences
 Patent: WO 0112659-A 802 22-FEB-2001;
 JOURNAL German Human Genome Project (DE)

FEATURES

source
 1..4574
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 1105 a 1166 c 1158 g 1145 t
 ORIGIN

Query Match 99.8%; Score 1487.8; DB 6; Length 4574;
 Best local Similarity 99.9%; Pred. No. 0;
 Matches 1489; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 atgagctggtctgtgggtgtgtgtatcccttggggctgtcttccgtgtcgtatcc 60
 DB 205 ATGAGCTGCTCTGTGGGTGTGTGTATCCCTTGGGCTGTCTGTCTGCGGATCC 264
 QY 61 caaggtacacctctgtcccaacgttcaactctctttagaagagctgtcagaatacag 120
 DB 265 CAAGGTACTCTCTGCCCAACGTCACTCTTGTAGAGAGCTGTCAAGAAATACAG 324
 QY 121 aacgagctcaactccgggttcgagagacatccccaaggaaggaagagatcttc 180
 DB 325 AACGAGTCTCACTCCCGGTCGCCAGAGCCATCCCAAGGAGACAAAGAGGATCT 384
 QY 181 atgtgtcaacaagaagcttggggccaggtgtcagcctcccaagcctccaactgtg 240
 DB 385 ATGTGTCAACAAGAAGCTTCGGGGCCAGGTGCACCTTCAGGCCCTCAAACTGAG 444
 QY 241 acctggatgacgaactgtggaagtcgtgtgacagcgttggcagtaatgtcgtggag 300
 DB 445 ACCTGGATGACGAAGTGTGAGAGTGTGTGACGCGTGGGCGAGTCAATCTGGAG 504
 QY 301 cagcggcccaacagctgtcgtgtgtgtatcagggcagaagccttggcgctcactgtg 360
 DB 505 CACGGGCCACACAGT 564
 QY 361 tctgctctccgggttccatgtgtcagcttctgtatgacgaaggtgaagagactac 420

Db	565	TATGGCTCTCCGGGGTTCCATGTCAGTCTCTGTAAGCAAGGTGAAGACTTAACCTTAC	624
Qy	421	ccctaccccgagagtgcaacccctctgtgtccagaaggctctcgggagctatgtgcacg	480
Db	625	CCCTAACCCGAGCAAGTCCAAACCCCTGGTGTCCAGAGAGTCTCGGGCCCTATGTGCACG	684
Qy	481	cactacacacaga taatttgggccacacaa caagctcgggtttgtcgtgtgaacacttc	540
Db	685	CACATACACAGATAGTTGGGCCAACCAACCAAGATCGGTGTGCTGTAAACCTGC	744
Qy	541	cggagaagtgtcgtctcggggagaagtttggagaa cgcggtctacttgtctgcattat	600
Db	745	CGGAAGTGTACTCTCTGGGGAAGAATTTGGGGAACGGCGGTCTACTTTGTCGCAATTAT	804
Qy	601	tctccaaaggagac tfgatttgaagaagccccctacaaagatgacggccctctctgag	660
Db	805	TTCTCAAAAGGGGAACGTGATTGGAGAAACCCCTTACAGAAATGGCCCGCTCTCTGAG	864
Qy	661	tgtccacccaagctatggaagcagctgcagaacaactgtgttaccgagagaaactac	720
Db	865	TGGCCACCCAGCTATGTGAGGACGCTGCAGGAACAATGTGTATCCGAGAAACCTAC	924
Qy	721	actccaaaacctggaagacgagacaggttgaatagtttggaagagcgtccactctgaaga	780
Db	925	ACTCCAAAACCTGTAAACGGACAGCAGATGATAGGTGAACGGCTCCACTCTCGAAGAA	984
Qy	781	aaccatglttggctccaaacggaaggttga tgaaccacacagcccaagaanaactctgag	840
Db	985	AACCATGTTTGGCTCCAAACCGAGGGTGTATGAGACCCAAAGCCCAAGAAACCTCTGCG	1044
Qy	841	gtcaactacatgacccaagtctgaagtgtgacaccaaagtgaagacaggtgtcaaggg	900
Db	1045	GTCAACTACATATACCCAAAGTGTGATGATGATACACCAAGATGAAGGACAGGTCAAAAGG	1104
Qy	901	tccacgtgtaaacaggtaccaagtgcacagcagcgtgccttgaaaccaaaagcgaagatcctt	960
Db	1105	TCCACGTTGAACAGGTACCAAGTCCGACAGGGCTGTGAAACCAAGCGGAAGATCTTT	1166
Qy	961	ggaagctgttcttaagaagaagctgtctagatatgcccgcgcgcacatccactaagagatc	1020
Db	1165	GGAACTCTGTTCTATGAAGAAGCTCTCTAGATATAGCCGCCCCCAATCACTACCGGATC	1224
Qy	1021	ctgagatgaacaaaggagcgtctgtgtgatatacacaagaaacgaggaagttccctctctgt	1080
Db	1225	CTGGATGTGCAAGAGGAGGCGCTGTGATATACCAAGGAAGGGAAGTCCCTCTTCTGTG	1284
Qy	1081	aagcttgagaaacagcggtgtagtccctcaagcaaatcaaaacttcaagatcattcatg	1140
Db	1285	AAGTGTGGAAGACACGGGGTGCAATCCCTCAGCAAAATCAAACTTCACGCTCATTCATG	1344
Qy	1141	gtgtcaaaagtgaagatgtgcaggaatttggacgtgtacaagacggtttgtcaagtgtggccg	1200
Db	1345	GTGTCAAAAAGTGAAGTGCAGAGATTTGGATGTAACACACCGTTTGCTCAGCTGTGCCG	1404
Qy	1201	tttgaagaagccaagaactcactgtcccaagaatccattgtctcggcacactgtgaagacgaa	1260
Db	1405	TTTGAAAAGCCAGCAACTCACTGCCCAAGATTCATATTGCCGACACATGCAAAAGACGA	1464
Qy	1261	ccttcctactctgggtctcgggtgttttggaaacaaatctatagagataactcaagatcttgc	1320
Db	1465	CTTTCTACTAGGGGCTCCGGGTGTTTGGAAACCAACATCTATGCAAGATTACTCAAGCATCTGC	1524
Qy	1321	aagcagctgtgcaacgcggaagatcatcaagcaacgagagttgaggtgtgacgttgacgtgtatg	1380
Db	1525	AAGACAGCCGTCGACGCGCGGAGTCAATCAGCAACGAGAGTGGGGGTGACGTGACTATG	1584
Qy	1381	cccggtgataaaagaagacactcgtgggtcgtctcaggaatgtgaagttcaagttgaagc	1440
Db	1585	CCCTGTGATTAAGAAAGAACCTTACGTGGGGCTGCTCAGGAATGTGAGTTCATCTGAAGC	1644
Qy	1441	ctggagctctctcgggatgtgaagagccttcgagatcttgtcgtgaagcag	1491

DB	1645	CTGGGAGTACTCTCGGAGATGGAAGCCCTTCGCGATCTTTGCTGTCAAGCAG	1695
RESULT	4		
LOCUS	HSB01829	4574 bp	linear
DEFINITION	Homo sapiens mRNA; cDNA DKFZp434B04.4 (from clone DKFZp434B04.4); complete cds.		
ACCESSION	AL136861		
VERSION	AL136861.1	GI:12053226	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	1 (bases 1 to 4574)		
TITLE	Mambull R., Heubner, D., Mewes, H.W., Gassenhuber, J. and Wiemann, S.		
JOURNAL	Direct Submission		
COMMENT	Submitted (15-JUL-2001) MIPS, Am Klopfersplitz 18a, D-82152 Martinsried, GERMANY		
	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de; consorrtium of the German Genome Project.		
	This clone (DKFZp434B04.4) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de. Further information about the clone and the sequencing project is available at http://mips.gsf.de/proj/cDNA/ .		
FEATURES	location/Qualifiers		
source	1..4574		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/map="333.4 CR from top of Chr16 linkage group"		
	/clone="DKFZp434B04.4"		
	/tissue_type="testis"		
	/clone_id="434 (synonym: htes3). Vector pSport1; host DH10B; sites NotI + SalI"		
	/dev_stage="adult"		
gene	205..1698		
	/gene="DKFZp434B04.4"		
CDS	205..1698		
	/note="Strong similarity to trypsin inhibitor."		
	/codon_start=1		
	/product="hypothetical protein"		
	/protein_id="CAB66795.1"		
	/db_xref="GI:12053227"		
	/db_xref="SPRMBL:Q9H0B8"		
	/translation="MSCVLGVPIPLGILFVCGSGQILPNTLLEILSLKYOHNSH SRVRAIPREDKEEILMLHNKLRGOVPGASNNEMYTWDELEKSAANAASQITWEG PSTLSIVSGIOLGAHWGRYSRPGHNSWYDEKDTYTPRSPCNMCSPERGSPMCT HYTLIWATNTKIGCAVNTCRKMTVMCEWENAVITVCNISPFGNNIGAPKNGKPC SECPSPGGSGCRNNLCRETEYTPRKEPTDMEINVEETAPLPEENHVALQPVMPKPC KTSVANVTWVRCIDTMKDRCKSGSCNRCVPCAGCGLNHAKLFTGLFYESSSSICRA AIHNGILDDGGGLVDITRNKGVPEFVKSERHGVSLSKYKSSSPFWSKVYKODIDCY TYVAQLCPFEKPAITHCRIRICPAHCKDEPSYAPVPGTNIYADTSLTINAHAGVIS NESGGVDVMPVPRKKTYVSLNNGVSSSLGPRGKAFRIFAVQ"		
	4540..4545		
	polya_signal		
	polya_site		
BASE COUNT	1105 a 1166 c 1158 g 1145 t		
ORIGIN			
Query Match	99.8%; Score 1487.8; DB 9; Length 4574;		
Best Local Similarity	99.9%; Pred. No. 0;		
Matches 1489; Conservative	0; Mismatches 2; Indels 0; Gaps 0;		
1	atgaagctgcgtcctggatggtatcacccttggagctgtcctgctgctgagatcc	60	
205	ATGAGCTGCCTCGGAGATGGAAGCCCTTCGCGATCTTTGCTGTCAAGCAG	264	
61	caagctaccctcctgcctcaacgctaccctctttagagagctgtcctcaagcaataccagc	120	

```

Db 265 CAAGGCTACTCTGCGCAACGTCACCTCTTATGAGAGAGCTGCTCAGCAATAACAGCAC 324
Qy 121 aacagatctactccgggttcgacagacatccccaaggaggaacaaggagatcttc 180
Db 325 AACGAGTCTACCTCCGGGTCGCCAGAGCCATCCAGAGGAGACAGAGGAGATCTTC 384
Qy 181 atctctcaacaagaagctctgggccaagtgcaagcctcccaacttgagtaacatg 240
Db 385 ATGCTGCACAAACAGCTTCGGGGCCAGTGCACCTCAGGCTCCCAACATGAGTACATG 444
Qy 241 acctggatgacgaacttgagaaagctctgctcagcgctgggccaagtgcaatctggag 300
Db 445 ACCGAGGATGACGAACCTGAGAGATGCTGCTGACAGCGGGGCCAGTCACTGATCTGGAG 504
Qy 301 caagggcccccacagctcgtcgtgtcatatcgagcaaacctgggcccctcaactgggcaag 360
Db 505 CACGGGCCCAACAGTCTGCTGTCATCGGCGAGAACTGGGGCGCTCACTGGGGCAGG 564
Qy 361 tatcgctccggggttcacatgtgcaagctcgtgtatgacagagtgaaagatacaactac 420
Db 565 TATGGCTCTCGGGGTTCCATGTGCANCTCTGTATGACAGGTGAAGGATACACTAC 624
Qy 421 cccctaccggagcgagtgcaacccctggctgcagagagagtgctcggggacctatgtcag 480
Db 625 CCTTACCCGAGCGAGTGCACACCCCTGGTGTCCAGAGAGTGTCTGGGGCTTATGTGCAGC 684
Qy 481 cactacaacagatagtttgggccacacacacacacacacacacacacacacacacacac 540
Db 685 CACTACACACAGATAGTTTGGCCACCAACCAACAGATCGGTGTCTGTGTGAACACTTGC 744
Qy 541 cgaagaatgacgtctcgtggggaagatttgagagacacggttctacttctgtcgaattat 600
Db 745 CGGAAGATGATGCTCTGGGGGAAAGTTTGGAGAAACGGGGCTTACTTGTGTCAATAT 804
Qy 601 tctccaaagggaactggaatttgagagagcccccatacaagaattggccggtcgtctcgag 660
Db 805 TCTCCAAAGGGGAACTGGATTTGGAGAAACCCCTTCAAGAAATGGCCGGCTGCTGTGAG 864
Qy 661 tggccacacagctatlgagagcagcttgacagaaacactgtgttctccgagaaacactac 720
Db 865 TGCCACCCACTATGAGAGGACGCTGCAGAAACAACTGTGTACCGAGAAACCACTAC 924
Qy 721 actccaaacactgaaacggagcagatgaaatgagtgagtgaaacagcgtccactcctgaag 780
Db 925 ACTCCAAACCTGAAACGAGACGAGATGATGAGTGGGAAAGGCTCCATTCCTGAAATA 984
Qy 781 aacctgtttgtctcaacacgaggtgagtgagacccacacacacacacacacacacacac 840
Db 985 AACCAATGTTGGCTCCAAACCGAGGGTGTAGAGACCCACCAAGCCCAACAAACCTCTGCG 1044
Qy 841 gtcaactacatgacccaagctcgtcagatgtaacacaaagatgaaagagtgcaaaagg 900
Db 1045 GTCAACTATCATGCCCAAGTGTGTAGATGTGACACCAAGATGAAGACAGATGCAAAAGG 1104
Qy 901 tccacgltgaacagtgtaaccagtgcaccagcaggctcgtcgaacacaaaggcgaatctt 960
Db 1105 TCCACGTTGTAACAGGTACAGTGCACAGACAGGCTCTGTAACCAAGCGGAAGATCTTT 1164
Qy 961 ggaagctgtttatgaaagctcgtctagcatatgcccgcgcacatccacacagagatc 1020
Db 1165 GGAAGCTGTCTTATGAAAGCTGTCTTACATATGCGCGCGCATCTACACAGGATC 1224
Qy 1021 ctgagatgaagaaggagcctggatgatalacacacaggaacggaaaggtccctctcgtg 1080
Db 1225 CTGAGATGAAGAAGGAGCGCTGGGATATACACAGGAAGGAGAGGTCCCTTCTTCGTG 1284
Qy 1081 aagctgagagagacagcggtgtgacgtccctcagcaaatataaaactccagctcatcag 1140
Db 1285 AAGTCTGAGAGACACGGCGTGCAGTCCCTACAGAAATTAACAACTTCCAGCTCATCATG 1344
Qy 1141 gtctcaaaagtgaagtgcaaggtttggaactgtacacagcgttgctcagctgtgccc 1200

```

```

Db 1345 GTGTCAAAAGTGAAGTCCAGGATTTTGACATGCTACACAGCCGTTGGCTCAGCTGTGCCCG 1404
Qy 1201 ttgaaagccagcaactactgcccagaatccatgtgtccgacactgcaaaagca 1260
Db 1405 TTTGAAAAGCCAGCAACCTACAGTCCCAAGAAATCCATGTCTCCGGCACACTGCAAGACGAA 1464
Qy 1261 cctccactagtggtcgggtgtttggaacaaacatactatgagatactcaagatctgc 1320
Db 1465 CTTCTACTGGGGCTCCGGTGTGTAACCAACATCTGTGAGATCTCTCAAGCATCTGTC 1524
Qy 1321 aagacagcgtgtgacggggaatcatcagcaacagagatgggggtgagtgagtgatg 1380
Db 1525 AAGACACCGCTGACGCGGGAGTCTATCAGCAACGAGATGGGGGTGACCTGGAGTATG 1584
Qy 1381 cccgtgataaaagaagaactacgtgggtcgtcgtcgaagatgagtgatcgtgaaagc 1440
Db 1585 CCGGTGATTAAGAAAGACCTACGTGGGCTGCTCAGGAATGGAGTTCACTGTAAAGC 1644
Qy 1441 ctggggaactcctcggatgtaagagccttcggatcttgcgtcagagcag 1491
Db 1645 CTGGGACTCTTCGGGATGAAAGGCTTCGGATCTTGTGTGACAGGCGAG 1695

RESULT 5
AX285067/c 4877 bp DNA linear PAT 20-NOV-2001
LOCUS Sequence 872 from Patent WO01/9556.
DEFINITION AX285067
ACCESSION AX285067
VERSION AX285067.1 GI:17045755
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (sites)
AUTHORS Lilie,J., Brown,J.L., Bolt,A. and van Hufel,C.
TITLE Novel genes, compositions and methods for the identification,
assessment, prevention, and therapy of human cancers
JOURNAL Patent: WO 01/9556-A 872 25-OCT-2001;
Millennium Predictive Medicine, Inc. (US)
FEATURES
source
BASE COUNT 1183 a 1201 c 1236 g 1257 t
ORIGIN
Query Match 99.8%; Score 1487.8; DB 6; Length 4877;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1489; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 atgagctgctgctcgtggatgltgcatcccttgaggctgctgtctctgtcgtcgatcc 60
Db 4657 ATGAGCTGCGCTCGTGGGTGTGTATCCCTTGGGGCTGTGTCTCGTCTCGGATCC 4598
Qy 61 caaggttaactccctggcccaacagtgactctttagaggagctgctcagcaataaccagcac 120
Db 4597 CAAGGCTACCTCTGCCCCAACGTCTCTTGAAGAGAGTGTCTGCAAGAAATACAGACAC 4538
Qy 121 aacagatctactccgggttcgacagacatccccaaggaggaacaaggagatcttc 180
Db 4537 AACGAGTCTACTCCGGGTCGCCAGAGCCATCCCAAGGAGGAGCAAGAGAGATCTCTC 4478
Qy 181 atgctgcaacaagaactctgggccaagtgtaagctcgaagctcacaatgagatgatatg 240
Db 4477 ATGCTGCACAAACAGCTTCGGGGCCAGGTGCAAGCTCAAGCTCAACATGAGATACATG 4418
Qy 241 acctggatgacgaacttgagaaagctctgctcagcgctgggccaagtgcaatctggag 300
Db 4417 ACCTGGATGACGAACCTGAGAAAGTCTCTGACAGGTGTGGCCAGTGCATCTGGGAG 4358
Qy 301 caagggcccccacagctcgtcgtgtcatatcgagcaaacctgggcccctcaactgggcaag 360

```


Db 3877 AACGATGTTGGCTCCACCGAGGCTGATGAGACCCACCACCAAGCAAAACCTCTCG 3818
 Qy 841 gcaactacatgacccaagctgcatgagatgagcaccagaatgagaagatgacaaagg 900
 Db 3817 GTCAACTACATACCAACCAAGCTGCTAGATGTGACCAAGATGAGAGAGAGTCAAGG 3758
 Qy 901 tccacgctgaacaggttacacagtcacagcagcagcagcagcagcagcagcagcagc 960
 Db 3757 TCCAGCTGTAACAGCTACAGTCCGACAGAGCTCCGTAACCAAGAGAGATCTTT 3698
 Qy 961 ggaagctgcttctatgaaagctcgtctagacatagccgagccgacatccacagagatc 1020
 Db 3697 GGAAGCTGTTCTATGAAGCTCTGATGACATATCCGCGCCGATCCATCAGGAGATC 3638
 Qy 1021 ctgagtgacaaaggagagcagctgtagatataccacaggaacaggaagctccctctctg 1080
 Db 3637 CTGATGACAAAGGAGGAGCTGATGATACACAGAGAGAGAGAGAGAGAGAGAGAG 3578
 Qy 1081 aagctgagagacacagcagctgacgtccctacagcaataacacacacacacacacac 1140
 Db 3577 AAGTCTGAGAGACACAGGCTGATGCTCCAGCAAAATCAAAACCTTCAGCTCATTCATG 3518
 Qy 1141 gctcaaaagtgaagtgacagatgttgaagctgacacagacacacacacacacacac 1200
 Db 3517 GTGCTAAAGTGAAGTACAGAGATTTGAGCTGACACAGAGAGAGAGAGAGAGAGAG 3458
 Qy 1201 ttgcaaaagcagaac 1260
 Db 3457 TTTGAAAAGCCAGAC 3398
 Qy 1261 ccttcacagtgctcagcagctgctgagacacacacacacacacacacacacacacac 1320
 Db 3397 CCTTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3338
 Qy 1321 aagaacagctgacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1380
 Db 3337 AAGACAGCGCTGACAGCGGAGGATATGACACAGAGAGAGAGAGAGAGAGAGAG 3278
 Qy 1381 cccgtgataaagaagaac 1440
 Db 3277 CCCGTGATTAAG 3218
 Qy 1441 ctgagagcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1491
 Db 3217 CTGGGAGACTCTCGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3167
 RESULT 8
 AX235369 2400 bp DNA linear PAT 11-SEP-2001
 LOCUS AX235369 Sequence 16 from Patent WO0162928.
 DEFINITION AX235369
 ACCESSION AX235369
 VERSION AX235369.1 GI:15593901
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 2400)
 Vernet,C.A., Fernandes,E., Shinkels,R.A., Macdougall,J. and
 Spaderna,S.K.
 Polypeptides and nucleic acids encoding same
 Patent: WO 0162928-A 16 30-AUG-2001;
 Curagen Corporation (US)
 FEATURES
 Location/Qualifiers
 1..2400
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 206..1699
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC69818.1"

/db_xref="GI:15593902"
 /translation="MSCVLGVYIPLILPLVRSOGYLLPNVTLLELLSKYOHNEH
 SRRRAIPREDKEELIIMLNKLJGQVOPASNMKEWTMDLEKSAAMASQITHEG
 PGLIVSICONLGAHMGRRSPDFHVSQSDYRDKYTPYPSQPMPCBERGCMCT
 HYQIIVATTNKLGCANVCRKMTWGEFWEANAYFVNCYSRKNMIGBAPKNGRPC
 SECPSYGSCKRNLCYREETTPREDTBNNEVERAPLPEENHVLDPVVRPTPK
 KTSAVNMYQVVRCDTKMKDRCKGSDCNRYQCPCGLNMRKAFISLFESSSTCR
 ALHIGLIDRGGSLVDITRNGKVFYFVSRHGVOSTLSKYPSSTSVASKVKYODDY
 TTVAOICPEKPAVHCPRJHCPAHCKDEPSYWAPEVNTIYADTSISIKTVAHGVIS
 NMSGGDVMPYDKKTKYVGSIRNGVOSLSLTPRDKAGAFRIEAVRQ"
 BASE COUNT 556 a 646 c 684 g 514 t
 ORIGIN
 Query Match 99.7%; Score 1486.2; DB 6; Length 2400;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1488; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 atgagctgagctcctggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 60
 Db 206 ATGAGCTGCTCTCTGAGGTGAGTATCCCTTGGGCTCTCTCTCTCTCTCTCTCTCTCT 265
 Qy 61 caaggtac 120
 Db 266 CAAGGCTACCTCTCTGCAACGTCATCTTAGAGAGCTGCTCAGCAATACACACAC 325
 Qy 121 aagagctcactcccggttcgacagacacacacacacacacacacacacacacacacac 180
 Db 326 AAGGAGCTCTACCTCCGGGTCCGACAGACATCCCGAGGAGACAGAGAGAGATCTCTC 385
 Qy 181 atgctcac 240
 Db 386 ATGCTCACACACACAGCTTGGGGCCAGTGACCTCAGGCTCCACACATGAGATCATG 445
 Qy 241 acctggtatgacgaactggaagaatctgctgacagctggtggtggtggtggtggtggt 300
 Db 446 ACCGTGGATGAGAGATGAGAGAGATCTGTCAGAGCTGGGCCACAGATGATGAGAG 505
 Qy 301 caaggtccac 360
 Db 506 CAGGCGCCACCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 565
 Qy 361 tatcgctccggttccatgltgacgtcctgtaagagagtgaaagacacacacacacacac 420
 Db 566 TATCGCTCCGGGTTCCATGTCAGTCTGTGTGACAGAGTGAAGACTACACCTAC 625
 Qy 421 cctaccgagagagtgac 480
 Db 626 CCTTACCGAGGAGTGCACACCTCTGTTGTCAGAGAGTCTCTGGGGCTATATGTGACG 685
 Qy 481 cactac 540
 Db 686 CACTACACACACATAGTTTGGGCCACACCAAGATCGGTTGTGCTGTAAACACCTGC 745
 Qy 541 cggagagatgactgctggtggtggtggtggtggtggtggtggtggtggtggtggtggt 600
 Db 746 CGAAGATGACTGCTGGGGAGAGTTTGGGAGAACCGGCTCTACTTGTGTCGAATAT 805
 Qy 601 tctccaaagggaactggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 660
 Db 806 TCTCCAAAGGGGAGACTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 865
 Qy 861 tgcac 920
 Db 866 TGCCACCCACACTATGAGAGCAGCTGAGAGAACACTTGTGTACCGAAGAACACTTAC 925
 Qy 921 actccaaactgaaagagacagagatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 980
 Db 926 ACTCCAAACCTGAAAGAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 985
 Qy 981 aacatgttggctcacaacgaggtgatgagacacacacacacacacacacacacacacac 1040
 Db 986 AACCATGTTGGCTCAACGAGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1045

RESULT	12
AX235367	
LOCUS	AX235367
DEFINITION	Sequence 14 from Patent WO0162928.
ACCESSION	AX235367
VERSION	AX235367.1 GI:15593899
KEYWORDS	.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 2305) Vernel,C.A., Fernandes,E., Shinkets,R.A., MacDougall,J. and Spaderna,S.K.
TITLE	Polypeptides and nucleic acids encoding same
JOURNAL	Patent: WO 0162928-A 14 30-ADG-2001;
FEATURES	Curagen Corporation (US)
source	Location/Qualifiers 1..2305

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
453..1604
/note="unnamed protein product"
/codon_start=1
/protein_id="C6C69817.1"
/db_xref="GI:15593900"
/translation="MTMGRTSRSPGFHVQSWYDEVKDYTPYPEECNPGCEKSGCP
CHTYIIVATNTKIKICAVNTCRKKTWGEVNAENAVFCVSPKGMTEAEYKRGK
PCESEPVYSGSCRNLIKRELYTPKPEIDBMEVFETAPLPENHHWLOPRVARTK
PKTSAPVNTQVIVRODPIKMRDKRCSGNCNQCQDAGCLNKKATIEPTEHSESSIC
RAAIHGLIDDKGGLVDITRNKRYPFYFKSRHRSVQSLSKYKSPSSPMWSKAVQD
CYTVYALDPFEKPAITHCRIPAHCKIDSESVYAPVFGNFIADTSLIKTEAVHAGV
ISNEGGDVLPEVDKKTYVSLNLQVSSLSLTCPDNGAFRTFAVRQ"

```

BASE COUNT	542 a	619 c	647 g	497 t	ORIGIN
Query Match	86.3%;	Score 1286.2;	DB 6;	Length 2305;	
Best Local Similarity	93.4%;	Pred. No. 4,5e-297;			
Matches 1393;	Conservative	0;	Mismatches 3;	Indels 99;	Gaps 1.
1	atgagctgcgtccctgggtgtgtgtcattccctcttgggtctgtctgtctgtctgtctgtcgtatcc	60			
206	ATGAGCTGCGGCTCGGGGTGTGTATCCCTTCGCGGCTGCTCTTCCTTCGCTCCGGATCC	265			
QY	61	caaagctacctctctgcccacgctcactctctttagagagatgtcttagcaataccaagc	120		
DB	266	CAAAGCTACCTCCGCGCCAACTCATCTCTTTAGAGAGACTGCTTCAGCAATAACACGAC	325		
QY	121	aaagagcttaactcccggtctcgacagagccatccccaaggagacaaaggagagatactctc	180		
DB	326	AAGAGTCTACTCCGCGGTCCCAAGAGCATCCCGAGGAGGACAAAGAGAGATTCCTC	385		
QY	181	atgctgcacaacaagcttcggggcagaggtgcagccctcaagcctccaacttgaatgatactg	240		
DB	386	ATGCTGCACAAACAGCTTCGGGCCAGGTGTCAAGCTTCAGGCTCCAACTGGAGTACATG	445		
QY	241	acctggagatgaacgaactggagaagtcctctgcagcgttggccagtcagtcagtcgtggag	300		
DB	446	ACCTGGAGATGACGA-----	459		
QY	301	caaggcccaacgaagctgtgtgtgtcattcgcgtgcagaaacttggcgtcacttgggcagg	360		
DB	460	-----	470		
QY	361	tatcgtctccgggttcattgtacgtctgtgtatgtgacagagtgtaagagatacaactac	420		
DB	471	TATCGCTCCGGGTTTCATGTGCAGTCTGTGTATGACAGGTGAAGGACTTACACCTAC	530		
QY	421	ccctaccgaagagatgtgaacccctcgtgtgtctcagaaggctcctcgggcctatgtgcag	480		
DB	531	CCCTACCCGAGGAGAGCAACCCGTGTGTCCAGAGAGTCTCGGGCCATGTGTGACG	590		
QY	481	caactacacagatagtttgggcacacacaaagaatcgggttctgtgtgtataaactgc	540		
DB	591	CACATCACACAGATAGTTGGGCGCACACCAACAAAGATCGGTTGTGTGTATACACTGC	650		
QY	541	cggaagatgactgtcttggggagaaagtttgggagaaacggtctactttgtctgcaattac	600		
DB	651	CGGAAGATGACTGTGTGGGGAGAAATTTGGGAGAACCGGCTCTACTTGTCTGCATTTAT	710		
QY	601	tctccaaggggaaacttgatctggagaagcccccttacaagaatgtgcggccctgtctgag	660		
DB	711	TCTCCAAAGGGGAACCTGGATTGTGANAACCCCCCTTACAAAGATGGCCCGCCTGCTGAG	770		
QY	661	tggccaaccagctatgtgagcgcagctgcagaacaacttgttlaacgagagaagaacctac	720		
DB	771	TGCCACCCACACTTGTGGAGCGAGCTGCAGAACAACTTGTGTACGAGAAAGAAACCTAC	830		
QY	721	actccaanaacttgaacagcagcagatgaatagagtgaaacggtcccatctctgaaga	780		
DB	831	ACTCCAAACCTGAAACGAGACAGATGAATGAAGTGGAAACCGCTCCCATCTCTGAAGAA	890		
QY	781	aaccatgtttggtctccaacgaggtgtatgagaccaccaagcccaagaanaacctctgcg	840		
DB	891	AACCATGTTTGGCTCCAAACGAGGGTGTATGAGACCCCAAGGCCCAAGAAACCTCTGGC	950		
QY	841	gtcaactacatgaccccaagtcgttaagttgtgacaaccaagaatgaagacaggttcaaaagg	900		
DB	951	GTCACATACATGACCAAGTCGTAGATTGTGAACCAAGATGAAGGACAGGTGCAAAAGG	1010		
QY	901	tccaagtataacagggtaccaggtcccgacgaggtgtgcttaaccaacaaggcgaagatctt	960		
DB	1011	TCCACGTTGTAAACAGGTATCCAGTCCCAAGCAGCTGCTCTAACCACCAAGCGGAAGATCTTT	1070		
QY	961	ggaagtcgttctatgaagaagctcgtctagcatatgtccgcgcgcacatccaactacaggtatc	1020		

Db 1071 GGAACCTGTTCTATGAAAGCTCTCTAGCATATGCGCGCCGCCCATCTACGAGATC 1130
 Qy 1021 ctggatgacaaaggaggagcctgttgatataccaggaaaggagccctctctgtg 1080
 Db 1131 CTGATGACAAAGGAGGAGCTGTGTGATATCCAGGAAGGAGAGTCCCTTCTTGTG 1190
 Qy 1081 aagctctgagagacagagcgtgagtcctcctaagaataacacacacacacacacac 1140
 Db 1191 AAGTCTAGAGACACGCGCTGACAGTCCCTGAGCAATATCAACCTTCAGCTATCATG 1250
 Qy 1141 gtgcacaaagtgaagtgcagagatttgactgtctacacagccctgtctacgtctgcg 1200
 Db 1251 GTGCAAAAGTGAAGTGCAGAGATTGTGACTGACACGACCGTGTCTGACGTCGCCG 1310
 Qy 1201 ttgaaaaagcgaac 1260
 Db 1311 TTTGAAAAAGCAGCACTACTGCGCCCAAGATCATGTCTCGGACACACTCAAAAGACGA 1370
 Qy 1261 ccttcctactgagcctcgggtgttttgaaacacacacacacacacacacacacac 1320
 Db 1371 CTTCTACTAGGCTGCTCGGCTGTTTGAACCAATCATCTATCACTCAACATCTGTC 1430
 Qy 1321 aagacagctgtgacgcgagagatcatcaagacagagagtggtgagcgtgagcgtgatg 1380
 Db 1431 AAGACACCGCTGCACGCGGAGATCATCAGCAAGAGAGTGGGCTGACGTGACGTGATG 1490
 Qy 1381 cccgtgataaaagaagac 1440
 Db 1491 CCGGTGATTAAGAAAGACCTACGTGGGCTCCTCAGGAATGAGTTCAGTCTGAAGC 1550
 Qy 1441 ctggagacctcctgagatgaagagcctcgcagatcttctgtctcagcagcag 1491
 Db 1551 CTGGGACCTCTCGGATGGAAGAGCGCTTCGGATCTTGTGCTCAGGAG 1601

RESULT 13
 AK027395 1690 bp mRNA linear PRI 15-MAY-2001
 LOCUS Homo sapiens CDNA FLJ14489 f1s, clone MAMMA1002881, weakly similar
 DEFINITION to GLIOMA PATHOGENESIS-RELATED PROTEIN.
 ACCESSION AK027395
 VERSION AK027395.1 GI:14042039
 KEYWORDS oligo capping; f1s (full insert sequence).
 SOURCE Homo sapiens mammary gland CDNA to mRNA, clone LIB:MAMMA1
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (sites)
 AUTHORS Isogai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
 Nishikawa, T., Nagai, K., Sugano, S., Aotsuka, S., Yoshikawa, Y.,
 Matsunawa, H., Ishii, S., Kawai, Y., Salto, K., Yamamoto, J.,
 Wakamatsu, A., Nakamura, Y., Nagahara, K., Masuho, Y. and Sasaki, N.
 TITLE NEDD human cDNA sequencing project
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1690)
 AUTHORS Isogai, T. and Otsuki, T.
 TITLE Direct Submission
 JOURNAL Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
 Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3951, Fax: 81-438-52-3952)
 COMMENT NEDD human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan: CDNA full insert sequencing;
 5'- & 3'-end one pass sequencing and clone selection; Helix
 Research Institute (supported by Japan Key Technology Center etc.)
 and Department of Virology, Institute of Medical Science,
 University of Tokyo.
 FEATURES
 source 1. 1690
 location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

CDS
 192..1307
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAB55081.1"
 /db_xref="GI:14042040"
 /translation="MSCVIGGVIPGLGLFLVCGSGYLLPNVTLLELLSKYOHNEH
 SRRRAIPREDKEELIHLNKRIGVQPOASNMEDLDEENSAAMSOCTIWEIG
 PTLISLIGONIGAHMGRVRSRSPGVQSVEDVDYRYPSPBCNPRECSPPKCT
 HTLOIVMTNKGICAVNTCKRMVWGSVWENAVFYCNISPKGMVIGEAYKGRPC
 SECPSPYSGSCNNICLYREYTPPEPDEENVEVATIPENHVMLOPRMPTPK
 KTSANVNTQVVRCDTKMKDKSCNRKYCCPAGCCLNHRKIKIGTLFESSSSICRA
 AHHGLIDDKGLVDITNGKVPFVKSEHGVSLR"
 BASE COUNT 410 a 469 c 489 g 322 t
 ORIGIN

Query Match 74.4%; Score 1109.6; DB 9; Length 1690;
 Best Local Similarity 99.6%; Pred. No. 7.8e-255;
 Matches 1112; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 atgagctgctcctggtgtgtgtatcccttgaggctcgtcttctgtctcgtgagatcc 60
 192 ATGAGCTGCTCCTCGGTGTGTGTATCCCTTGAGGCTGCTTCTCTGCTGCGGATCC 251
 Qy 61 caagctactctctgcccacagtcactctctagaagagctgtctcagaataacacagac 120
 Db 252 CAAGCTACTCTCTCCCAACCTCCTCTTAGAGAGCTGTCTCAGCAAAATACAGAC 311
 Qy 121 aacgagctcactcccggttcgagagacatccccaagagagagagagagagatctc 180
 Db 312 AACGAGTCTCATCTCCGGGCTCCGAGACCATCCCAAGAGAGACAAAGAGATCTCTC 371
 Qy 181 atctgtcaacaagaagcttcggggccaggtgtgagcctcagagcctccaagatgagatc 240
 Db 372 ATGCTGCAACAAGACTTCGGGGCCAGGTGCAAGCTCAGGCTCCCAATATGATATG 431
 Qy 241 acctgagctgcgaacttgagaagctctgctcagcgtgtggccagtcagtcagtcagc 300
 Db 432 ACCTGGAATGACGAAGCTGGAAGATCTGCTCAGCTGGGCCAATCATCATCTGGGAG 491
 Qy 301 caagggccacagctctgtgtctcattcgggagagaaacctgggcctcacttggggcag 360
 Db 492 CAGGGCCACACAGCTCTGCTGTGTCTGTCATCGGCGAAGAACTGGGCGCTCACTGGGCA 551
 Qy 361 tatcgtctcgggggttcattgtgcagtcctgtgatacagagtgagagactacactac 420
 Db 552 TATCGCTCTCCGGGGTTCATGTGCAGTCTGATATGACGAGTGAAGACTACACTAC 611
 Qy 421 ccttaccggagcggatgaaccccttggttcagaagagtgctcggggcctatgtcagc 480
 Db 612 CCTTACCGGACGAGTGCACCCCTGGGTCTCAAGAGAGTCTCAGGGCCTATCTGACG 671
 Qy 481 cactacacacagatagtttggtccacacacacacacacacacacacacacacacacac 540
 Db 672 CACTACACACAGATAGTTTGGGCCACACACACAAAGTCGTTGTCTGTGAACCTGTC 731
 Qy 541 cggagaagatgctctggtggagaagatttgagagaagcggctactacttctgtcgaatt 600
 Db 732 CGGAAGATGACTGTCTGGGGAGAGTTTGGAGAACCGGCTCTACTTGTCTGCAATTAT 791
 Qy 601 tctcacaagggagctgatttgagagagcccccacagaagaatggccgctctctctgag 660
 Db 792 TCTCCAAAGGGAGCTGATTGGAGAACCCCTTACAGATAGGCGCGCTCTCTCTGAG 851
 Qy 661 tgcacccacagctatggagcagctgcaagaaacacttgttaccgagagaacactac 720
 Db 852 TGCCACCCACAGCTATGAGGACAGCTGCAAGAACAACTTGTGTACGAGAGAAACTTAC 911
 Qy 721 actccaacactgaagaagcagatgaaatgagtgagtggaacagcgtccactctcagaaga 780

|||||
 Db 912 ACTCCAAACCTTAAACGAGATGAAGAGTGAAGAGGCTCCATCTCTGAGAGA 971
 QY 781 aaccatgttggctccaccgaggtgatagagccaccgaagcccaagaacccctctg 840
 Db 972 AACCATGTGGCTCCACCGAGGTGATGAGACCCACGAGCCCAAGAAACCTCTGGG 1031
 QY 841 gtcaactacatgaccccaagtcgtcagatgtgacacaaagtaaggaaggtgcaagg 900
 Db 1032 GTCAAACTACATGACCCAAAGTCGTACATGTGACACCAAGATGAGACAGTCAAGGG 1091
 QY 901 tcaacgtgtaacaggtacccaggtccagaggtctgctgtaacacaaaggaagatctt 960
 Db 1092 TCCACGCTTACAGGTACAGTCCAGTCCAGGCTGCTGACACCAAGGGAAGATCTTT 1151
 QY 961 ggaagctgtctatagaagctcgtctagatalgacgcgcgcacacacacacagggatc 1020
 Db 1152 GGAACCTGTCTTATGAAACCTGCTAGCATATGCGCGCCGCACTACGGGATC 1211
 QY 1021 ctgagatgacaaagagagctcgtgtgatatcaacaggaaggaaggtccctctctgtg 1080
 Db 1212 CTGATGATACAAAGGAGGCTGTGTGATATCACAGGAACGGGAAGTCCCTCTCTG 1271
 QY 1081 aagctgagagacacagcgtgacgtccctcagcaaa 1116
 Db 1272 AAGTCTGAGACACAGGCGCTGCACTCCCTCAGATTA 1307

RESULT 14
 AF109674 3052 bp mRNA linear ROD 27-SEP-2000
 LOCUS complete cds.

DEFINITION Rattus norvegicus late gestation lung protein 1 (Lg11) mRNA,
 complete cds.
 ACCESSION AF109674
 VERSION AF109674.1 GI:4324681
 KEYWORDS Norway rat.
 SOURCE Rattus norvegicus
 ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 3052)
 AUTHORS Kaplan, F., Ledoux, P., Kassamali, F.O., Gagnon, S., Post, M.,
 Koehler, D., Delmaling, D., and Swezey, N.B.
 TITLE A novel developmentally regulated gene in lung mesenchyme: homology
 to a tumor-derived trypsin inhibitor
 JOURNAL Am. J. Physiol. 276 (6), L1027-L1036 (1999)
 MEDLINE 99292450
 PUBMED 10362728

REFERENCE 2 (bases 1 to 3052)
 AUTHORS Kaplan, F. and Swezey, N.B.
 TITLE Direct Submission
 JOURNAL Submitted (26-NOV-1998) Human Genetics, McGill University- Montreal
 Children's Hospital Research Institute, 2300 Tupper Street,
 Montreal, QC H3H 1P3, Canada
 Location/Qualifiers

FEATURES
 source
 1..3052

gene
 /organism="Rattus norvegicus"
 /strain="Wistar"
 /db_xref="taxon:10116"
 /cell_type="fibroblast"
 /tissue_type="lung"
 /dev_stage="20 day old fetus"
 1..3052
 /gene="Lg11"

CDS
 /note="hormonally and developmentally regulated"
 265..831
 /gene="Lg11"

/note="Expressed in late gestation lung mesenchymal cells
 in rat, mouse, and human; similar to tumor-derived trypsin
 inhibitor"
 /codon_start=1
 /product="late gestation lung protein 1"

/protein_id="AA016986.1"
 /db_xref="GI:4324682"
 /translation="MLINKRGVYPPASMEYTWDELEERSAAMORCLMEHPA
 SLVLSIGONLAHEWGRYRSFPHQSDYDEKYYITPYPIECNPMPERSGSMCTHY
 TOMYMTATNKIGCAVHTCRSMWGDWIEVAIVLYCNVSPKGMWISAPYKHPQSE
 CPSSYGGGCRNNLCYREHRYHOKREHMR"
 BASE COUNT 695 a 815 c 851 g 691 t
 ORIGIN

Query Match 65.3%, Score 974.2; DB 10; Length 3052;
 Best Local Similarity 79.2%; Pred. No. 1.9e-22;
 Matches 1181; Conservative 0; Mismatches 308; Indels 2; Gaps 2;

QY 1 atgagctgctcctgtggtggtgcatcccttgaggcctgctcctgctgctgagatcc 60
 Db 86 ATGAGCTGCTCTGTGAACATATGGTCCCGGTGGGGCTGGCTCCGCTGTGTGGGAGTA 145
 QY 61 caagctacccctctgcccacgtcactctcttagagagctgtctagcaaaataccagac 120
 Db 146 CAGGCTTTTCTCTCCCAACACATGAGCCTGAGAGGCTGCTCAGCAAGTACAGCAGC 205
 QY 121 aacagatcactcccggtctcgagagacatccccagggaggaagagagagatctc 180
 Db 206 ACAGAGCCACACTCACGCTGTGGGAGGCGCATCCCATGTGCAGAC-CGTCAGAATCTTC 264
 QY 181 atgtctacaaacagcttcggggcagagtgacagctcagcctcaccacatgagatcag 240
 Db 265 ATGCTGACACAAAGTGGAGGTGAGTATCCCGCTGCCCAACATGAGTACATG 324
 QY 241 acctggatgacgaactgtgagaagctcgtcagcgtgtggtgagctgagctggag 300
 Db 325 ACCTGGATGAGAGAGCTGGAGAGTGGGAGCGGCTGGGACACAGAGGCTGTTGGAG 384
 QY 301 caaggcccccacacagctgtgtgtgtcctcagcagcaggaaccttgagcctgagcag 360
 Db 385 CATGGGCGCGCCACAGCCTTCTGTCATCGGGCAGAACCTGGCTGTGCACGGGGCAGG 444
 QY 361 tatcgtctccggggttcacatgtcagctcgtgtatgacaggtgaaagtaacacac 420
 Db 445 TACCGCTCTCTGGTTCACACGTGAGTGTGATGAGAGGTGAGGATTTACCTAC 504
 QY 421 ccttaccagagagtgagacccctgtgtcagagaagtgctcagggagctgtgacag 480
 Db 505 CCGTACCCCCACGAGGACCAACCATGTGCTCCGAGCGCTGCTCAGAGCACTGTGAC 564
 QY 481 cactacacacagatgttggggccaccacaaagaatcgtgtgtgtgtgtgacacctgc 540
 Db 565 CACTACACACAGATGTTGGTGGCCACACCAACAAATTTGGCTGTGCACACTGT 624
 QY 541 cggagaatgactgtctgggagaagtttgggagaacgcygltacttcttgcgaattat 600
 Db 625 CGGAGCATGTGAGCTCTGGGAGACATCTGGGAAMCGCGTGATCTGTGCAATTAT 684
 QY 601 ttcccaaggagagactgattgagaagccctcacaagaatgagcgcgcctgtgag 660
 Db 685 TCTCCAAAGGGAACCTGGATTTGGTAGGCTCCTTACAAACATGCGCGCCCTGCTGAG 744
 QY 661 tgcaccaccatctatgagagcagctgcaagaacaacttgttaccgagaagaactac 720
 Db 745 TGTCCGTCAGATGATGAGGTGCTGCGGAGCAACACTGTTATCCGAAAGAGCACTAC 804
 QY 721 actcaaaacctgaagagcagatgaaatgaaatgaaacggctccactcttgaaga 780
 Db 805 CATCAGAAACC-GGAGTGATGATGAACGAAGTGAATCACCCCTGCTCCCGGAA 863
 QY 781 aaccatgttggctccaccgaggtgatagacccacacgaagcccaagaacccctctg 840
 Db 864 ACCCATGTGTGGTTCAACCAAGGTGTCAAAOCCTCMAAAGACCAAGAGCCCAAGTC 923
 QY 841 gtcaactacatgaccccaagctgtcagatgtgacacaaagtaaggaaggtgcaagg 900
 Db 924 GTCAACTTCATGACCAAGGTGTCCACTGTGACACCAAGATGAAGGATCATGCAAGGA 983


```
OY 1163 atttgactgtacagaccgttgcagctgtgccgtttgaaaagccagcaactcact 1222
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1284 CTGTGACTTGTGAAACACTGTGTGACAGCTCTGTCCATTTCATTAAGCCTGCTTACATT 1343
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1223 gcccaagaatccatgtcccgacacatgcaagaagaaacccctccctactggtccggtgt 1282
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1344 GCCCAGAGGTATACGTGCTGTGTAAGTATGCAAGCAAAATCCACATTTATGCTGCTGTAA 1403
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1283 ttggaaccaacatctatgcaatacctcaagcatctgcaagaagctgtgcaagcggtgag 1342
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1404 TTGGAACTCGAGTTTATCTGTCTGACAGTATGTCAGAGCAGACAGTACATGCTGGAG 1463
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1343 tcatcaagcaacgagatggtggtgacgtgagcgtgagccgtgagataaaaagaagacct 1402
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1464 TGGTTCGAAATCA---CGGTGTTATGTGTATGTATGCTGTGTGACAAAAGAAAGACT 1520
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1403 acgttgggtcgtcaggaatgagttcagttcgaagaccttggggaactcctcgggattgaa 1462
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1521 ACATTGCTTCTTTTCAGAAATGGAATCTCTCAGAAAGTTTACAGAAATCTCTCAGAGGAA 1580
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1463 aggccttcgggactcttgcgtgt 1484
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1581 AGGCATTTCAGAGTGTGTGCTGT 1602
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: May 25, 2002, 13:22:02
Job time: 10318 sec

Tue May 28 16:10:02 2002

us-09-667-380a-1.rst

No suitable data
Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2002, 10:25:14 : Search time 2064.9 Seconds
(without alignments)
9745.731 Million cell updates/sec

Title: US-09-667-380a-1
1491
Perfect score: 1 atgagctgcgtctcgtggtg.....ggatcttcgtctcagcag 1491

Sequence: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Scoring table: 13736207 segs, 674847542 residues
27472414

Total number of hits satisfying chosen parameters:
Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: EST:
1: em_estba:*
2: em_estbm:*
3: em_estbn:*
4: em_estbu:*
5: em_estbv:*
6: em_estbw:*
7: em_estbx:*
8: em_estby:*
9: em_estbz:*
10: em_estc:*
11: em_estd:*
12: em_este:*
13: em_estf:*
14: em_estg:*
15: em_esth:*
16: em_esti:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	967	64.9	2929	AK019034	Mus musculus
2	509	34.1	651	BF361052	MR2-OT007
3	447.2	29.5	459	BF998318	PM1-GN018
4	439.8	29.1	680	BF998318	PM1-GN018
5	434.2	28.3	680	BF998318	PM1-GN018
6	421.4	27.6	748	BF185378	603382445
7	411.8	26.8	835	BF185378	603382445
8	401	26.8	835	BF185378	603382445
9	398.2	26.7	762	BF185378	603382445
10	398.6	26.7	762	BF185378	603382445
11	395	26.2	822	BF185378	603382445
12	391.2	26.0	783	BF185378	603382445
13	387.8	25.7	763	BF185378	603382445
14	383.6	25.4	763	BF185378	603382445
15	379.2	25.2	791	BF185378	603382445
16	375.8	24.1	573	BF185378	603382445
17	358.8	24.1	573	BF185378	603382445

RESULT	1	2929 bp	linear	HTC 19-JAN-2002
LOCUS	AK019034			
DEFINITION	Mus musculus 10 day old male pancreas cDNA, RIKEN full-length			
ACCESSION	AK019034			
VERSION	AK019034.1			
KEYWORDS	HTC, CAP trapper.			
SOURCE	Mus musculus (strain: C57BL/6J) 10 day old male pancreas cDNA to			
ORGANISM	Mus musculus			
REFERENCE	1 (sites)			
AUTHORS	Carninci, P., and Hayashizaki, Y.			
TITLE	High efficiency full-length cDNA cloning			
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)			
PUBMED	99279253			
REFERENCE	2 (sites)			
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,			
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to			
JOURNAL	prepare full-length cDNA libraries for rapid discovery of new genes			
PUBMED	Genome Res. 10 (10), 1617-1630 (2000)			
REFERENCE	3 (sites)			
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,			
TITLE	Kono, H., Akiyama, U., Nishi, K., Katsunuma, T., Tashiro, H., Itoh, M.,			
JOURNAL	Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,			
PUBMED	Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,			
REFERENCE	Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,			
AUTHORS	Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and Hayashizaki, Y.			
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format			
JOURNAL	sequencing pipeline with 384 multiplexed sequence			
PUBMED	Genome Res. 10 (11), 1757-1771 (2000)			

QY	1021	ctgatgtacaagaaggagcctgtgttgatataccaagaagaaggaggtccctcttcctgtg	1080
Db	1217	ATGCATGATCAGAGGTGGCCTGTGGCATGTCCACGAGAAATGGATGGTACCCCTCTTTGTTC	1276
QY	1081	aagctcagagaacacggcgctgcagctccctcagaacaatatacaaccttcacgtcatcatg	1140
Db	1277	AAGTCTCAGAAAATTTGGCATGTGATTCCTCTTGACAAATATACAAAGCCGATCTACTCTTCACT	1336
QY	1141	gtgtcaaaagtgaagatgtcaggaatttgaactgtctacacggacggttgttcagctgtgtcccg	1200
Db	1337	GTGTCAAAAGTGAACAGAGACGCCGCTGTGACTGCCACGCCACAGGTGCAGACAGCTGTGCCCC	1396
QY	1201	tttgaagaacgacgaactcactctgtcccaagaaatccattgttccggacacctgtcaaacgaa	1260
Db	1397	TTTCGAAAGCCGGCCACCCACTGCGCCGAGAAATCCAGTGTCTGCGGCTGTGGAGAGAG	1456
QY	1261	ccttctactatggtctcgggtttgtgaaaccaaatcatatgcagatlaacttaagaatctgc	1320
Db	1457	CCATCCTTATTGGGCTCCTGTGTATGGAAACAAACATCTATGTGATACCTTCCAGCATTTGT	1516
QY	1321	aagaacagctgtgcacgacggagatcaacgaacgaagtgtgggtgtgaactgtgaacgtgatg	1380
Db	1517	AAGCGCGCTGTACAGCGAGTGTCTATCTTTGACAGAGTGTGGGTATGACGATGTGATG	1576
QY	1381	cccgttgataaanaaagaacactcagttgggtctcgctcagaagatgtgaattcagttcgtgaagc	1440
Db	1577	CCCGTGGACAAAANAAGAGCTTACGTGTGGCTCCCTCAGAGAACGGGGGTCAAGTGGAGAGC	1636
QY	1441	ctggaggaacccctcggagatgtgaaagcccttcggatccttctgtctgttaaggcag	1491
Db	1637	CTGAACACTCTCTAAGAACGAAACGCTTCCGGAGATCTTGCCGTCAGGCAG	1687

[illegible]

Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel.: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/ICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?rl=MR2&t2=MR2-OT0079>)
230500-009-a08&t3=2000-05-23&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 561.
Location/Qualifiers
1..651
FEATURES
Source
/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone_lib="OF0079"
/der_strage="Adult"
note="Organ: ovary; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
150 a 179 c 161 g 161 t

```

[illegible]

	RESULT	3
LOCUS	BC011536	
DEFINITION	PM1-0N0180-111200-004-d10 GN0180 Homo sapiens cDNA, mRNA sequence.	
ACCESSION	BC011536	
VERSION	BC011536.1	
KEYWORDS	GI:12459833	
SOURCE	EST.	
ORGANISM	human.	
REFERENCE	Homo sapiens	
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Euthera; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 482)	
	Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordalo, S., Costa, F. F.,	

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

JOURNAL MEDLINE
COMMENT
20202663
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=PM1&t2=PM1-GN0180-
11100-004-d10&t3=2000-12-11&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 14
High quality sequence stop: 482.

FEATURES

Source

```
1. 482
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0180"
/dev_stage="Adult"
/note="Organ: Placenta, normal; Vector: puc18; Site_1: SmaI
; Site_2: SmaI; A mini-library was made by cloning
products derived from ORESTES PCR (U.S. Letters Patent
Application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the puc 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."
BASE COUNT
136 a 127 c 126 g 93 t
ORIGIN
```

Query Match 30.0%; Score 447.2; DB 10; Length 482;
Best Local Similarity 97.2%; Pred. No. 1.6e-103;
Matches 455; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```
QY 537 ctgcggaagatgactctctggggaagatttgggaacgcggtcttactgtctgca 596
DB 15 CTGCTCGAATATGACTGTCTGTAGAAATTGGAGAACCGCGCTACTTATGCTGCAA 74
QY 597 ttattctcaaaaggagactggattggaagacccctcaagaatggccgctgtctc 656
DB 75 TTAATTCGCAAAAGGGAACGTGATGGAAGCCCTTACAAAGATGCGCGCTGCTC 134
QY 657 tgggtgcacacagctatggagcagctctgaaagaacactgtgtaccggaagaac 716
DB 135 TGAGTCCCAACCCAGCTATGAGAGCAGCTCAGGAACAATTGCTTACCGAGAGAAC 194
QY 717 ctacactccaacacacggaagacagatgaaatggatggatggatggatggatggat 776
DB 195 CTACACTCCAAAACCTGAAACGAGACGAGATGAGTGAAGTGAAGTGAAGTGAAGTGA 254
QY 777 aaaaaacatgttggctcacaacgaggtgatgagaccacacacacacacacacac 836
DB 255 AAAAAACCATGTTTGGCTCCAAACCGAGGTGATGAGACCCACCAACCCCAAGAAAACCTC 314
QY 837 tgcggtcaactcatgacccaagtgcgtcagatgtgacacacaatgaaagacagtgacaa 896
DB 315 TCGCGTCAACTCATGATGACCAAGTGTGATGATGATGATGATGATGATGATGATGATGAT 374
QY 897 agggatcacatgttaacaggtacacagtgccagcagctcctgaacacacacacacacac 956
DB 375 AGGGTCAACAGTGTAAAGGTACAGTGTGACAGGCTGCTGAAACCAAGGCGCAAGAT 434
QY 957 ctgggaagctgttctatgaaagctcgtctagcatatgcccgcgcgc 1004
```

DB 435 CTTTGGAATCTGCTCTATGAAAGCTGCTTACGATATGCGCGCGCCG 482

RESULT 4
LOCUS BF998318/c 459 bp mRNA linear EST 24-JAN-2001
DEFINITION OVO-GN0212-101100-484-c03 GN0212 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF998318
VERSION BF998318.1 GI:12433634
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagal, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
JOURNAL MEDLINE
COMMENT
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=OV0&t2=OV0-GN0212-
101100-484-c03&t3=2000-11-10&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 22
High quality sequence stop: 456.

FEATURES

Source

```
1. 459
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0212"
/dev_stage="Adult"
/note="Organ: Placenta, normal; Vector: puc18; Site_1: SmaI
; Site_2: SmaI; A mini-library was made by cloning
products derived from ORESTES PCR (U.S. Letters Patent
Application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the puc 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."
BASE COUNT
89 a 128 c 122 g 120 t
ORIGIN
```

Query Match 29.5%; Score 439.8; DB 10; Length 459;
Best Local Similarity 99.3%; Pred. No. 1.2e-101;
Matches 452; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```
QY 1008 ccaactcggatctctgtatgacaaggagcctgtgtgatataccaggaagggaagt 1067
DB 459 CCAGTACGGGATCTCTGATGACAAAGGAGGCGTGTGATATACACAGACGGAAGGT 400
QY 1068 ccactcttcgtgaagtctgagagacacagcgctgcctccctaacaaatcaaaccttc 1127
DB 399 CCCTCTCTTGTGGAATCTGTGAGAGACACGCGGTGCGCTCCCTCAGCAAAATACAACTTC 340
QY 1128 cagctatcatcgtgtctcaaaagtgaagtgcaggaattggactgtctacagcagcttgc 1187
DB 339 CAGCTCATTCATGATGTTGCTCAAAAGTGAAGTGCAGGATTTAGACTGCTACACGACGTTGC 280
```

QY	1188	tcagcgtgtccggtttgaaagccagcaactcactcgtcccaagaatcaltgtccggaca	1247
Db	279	TCAGCTGTGCCCTTTGAAAGCAGCAACTCAGTCCCAACATCCATTGTCCGGACCA	220
QY	1248	ctggaagaagcaactcttcactcgtggtcgcgttgattgtgaaaccaactctatgtcagatac	1307
Db	219	CTGGAAGAGCAACCTTCTTACTGAGGCTCCGGGTGTGGACCAACATCTATGCGATAC	160
QY	1308	ctcaagcatctgcaagaagcagctgtgacgcgcgggaatcatcagcaacggagatgggggtgca	1367
Db	159	CTCAAGCATCTGCAAGACAGCAGCTGTGCACCGGGAGATCATCACACAGAGCGGGAGTGA	100
QY	1368	cgtggaogltatgcccgtgtgataaagaagacactcgtggtcgtcgtcgaagatlgagt	1427
Db	99	CGTGACCTGTATGCCCTGTGATATAAAGAGACCTACGTGGCGCTGCTCAGAAATGAGT	40
QY	1428	tcagtcctgaagccttgggagactcctcggatlgaa	1462
Db	39	TCAGTCTGAAGCCTGGGGG-CTCCTCGGATGGA	6
RESULT	5		
Bg175425		900 bp	mRNA
LOCUS			linear
DEFINITION			EST 06-FEB-2001
ACCESSION	Bg175425		
VERSION	Bg175425		
KEYWORDS	Bg175425.1	GI:12682128	
SOURCE	EST.		
ORGANISM	house mouse.		
REFERENCE	Mus musculus		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
JOURNAL	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
COMMENT	1 (bases 1 to 900)		
	NIH-MGC http://mgc.nci.nih.gov/ .		
	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: Gilbert Smith, Ph.D.		
	CDNA Library Preparation: Life Technologies, Inc.		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNL at:		
	http://image.llnl.gov		
	plate: LLM10262 row: p column: 19		
	High quality sequence stop: 686.		
	Location/Qualifiers		
	1..900		
	/organism="Mus musculus"		
	/strain="FVB/N"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:4460802"		
	/clone_lib="NCI-CGAP_Mam1"		
	/tissue_type="tumor, biopsy sample"		
	/dev_stage="10 months, virgin"		
	/db_host="DH10B"		
	/note="Organ: mammary; Vector: pCMV-Sport6; Site_1: salt;		
	Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.		
	Library constructed by Life Technologies. Investigator		
	providing samples: Gilbert Smith, NIH"		
BASE COUNT	175 a	248 c	199 t
ORIGIN			
Query Match	29.1%	Score 434.2;	DB 10; Length 900;
Best Local Similarity	80.7%	Pred. No. 4.4e-100;	
Matches 519; Conservative	0;	Mismatches 123;	Indels 1; Gaps 1.
849	catgaccgaagcgtcgtcaatgtgacaccaagaatgaagacaggtgtcaaaaggtccacggt		908

Db	Accession	Gene	Chromosome	Start	End	Strand	Length	EST	EST Date
Bb	10	CATGACCCAACTGGTCCACTGTGACACCAAGATGAAGGACTATGCAAGGATCCACGNG	69						
Oy	909	taacaggtaccagagtcaccaggaagctgtccctgaacacacaagggagaatctttggaagttc	968						
Bb	70	TAAAGGATACCAAGTCCACAGAGGCTGTCTGAGAACAAAGCCAAAGTTC-TTGGCTCTCT	128						
Oy	969	gtctctatgaagcttcgtcttagcatatccggccggccgcatcacaatacgaagatcccttgatga	1028						
Bb	129	GTATTATGAAAGTTCCTCCACCATATGCGAGCTGCAAGCTGCTATCTACTACGGTGTCATTCATGA	188						
Oy	1029	caagggaagccctgtgtgatactaccacaggaaggggaaggtccctctctgtgaagctga	1088						
Bb	189	TGAGAGTGCCCTGGTGGATGTCCACAGGAATGGATGGTACCCCTTCTTTTCAAGCTCA	248						
Oy	1089	gagacaagggcgtgtgaagtcctccatcgaaataataaaccttcagcgtcatcatctgtgtcaaa	1148						
Bb	249	GAAAAATGGCATGAGAGTCCCTGAGCAATACAAAGCCGCTTACCTCTTCACTGTGTCAAA	308						
Oy	1149	atgtgaagtgagcaggaatttggactgtctaaacagacgctgtctcaagctgtgcocgttga	1208						
Bb	309	AGTGAACAGAGAGCGCCGTGGACTGTGCACAGCCAGGTCCGACAGCTGTGCCCTTTCAGAA	368						
Oy	1209	gccagaactcactgcgccaaagatccatctgtccggaacacgtgcgaagagcggaacttccta	1268						
Bb	369	GCCGGCCACCCACTGGCCGACAGATCCAGTCTCTGCGCTGTGGAGAGAGCCATCTTA	428						
Oy	1269	ctgggtctcgggtgttttggagaccaaactcatgtcagataacctaaagcatctcgaagaacgc	1328						
Bb	429	TTGGGCTCTGTGTATGAGACCAACATCTATGCTGATTACTTCACACATTTGHAAGGCGCG	488						
Oy	1329	tgtgacgcgggagtgatcaagcaacagagagtgagggtgtgaagctgtgacgtgtgagccgtgtga	1388						
Bb	489	TGTGCACGCGAGGTGTCATCGTTGACAGAGTGTGGTGGCTTTGCGAGATGTGATGTCCTGTGA	548						
Oy	1389	taaaagaagaacctagctgtgtgtcgcgtcggacggaaatggatgcagctgaagccttggagac	1448						
Bb	549	CAAAAACAAAGCTACCTGGGCTCCCTCAGGAACGGGGGTGAGTGGGAAGGCTTAACAC	608						
Oy	1449	tcctcggatgtgaaagccttcgcggaatcttctgtctgaagcag	1491						
Bb	609	TCTCTCAGACGCGAAAGCGCTTCCGGATCTTTCGCGACAGCAG	651						

Tissue Procurement: Jeffrey Green M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 plate: LLMJ2058 row: 1 column: 12
 High quality sequence stop: 748.

```

/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="Dh10B"
/notes="Organ: mammary; Vector: pCMV-SPO6; Site_1: SALL1
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT
library constructed by Life Technologies. Investor:
Providing samples: Jeffrey Green, M.D., NIH"

```

[illegible]

406 CACGGGCCCCCGCCCTTCAGAGTCCATCGGCAACACTGGCTGTGCATCTGGGGCAG 465
361 TATGCTCTCCGGGATGCCATGTGCAGTCTCTGTATGACGAGGTAAAGACTAACCTAC 420
466 TACCGCTCTCCGGGTTCATGTGCAGTGCATGTGACGAGCTAAGGATTACACTAC 525
421 CCTACCCGAGCGAGTGCACCCCTGCTGTCCAGGAGTGTCTCGGGGCTATGTGACG 480
526 CCATACCCCGACAGTGCACCTCAGCGTCCCGGAGCCGCTGTCAAGCCCATGTGCACC 585
481 CACTACACCAATATGTTTGGGCAACCAACAAGATCTGCTGCGCCGTCACACCTGT 645
586 CACTACACCAATATGTTTGGGCAACCAACAAGATCTGCTGCGCCGTCACACCTGT 645
541 CCGAAGATGACTGTCTGGGAGAAATTTGGGAGAACGCGTCTACTGTCTGTAATAT 600
646 CCGAATCATGAAGTGTGGGGAGCACTTGGGGGAATGCCGCTGTTTCCTGCTGCANATAT 705

Db 706 TCTCCCAAGGAACTGATTGGCGAGCCCTCTACAGAT 748

RESULT 8 835 bp mRNA linear EST 31-0CT-2000
 BF181843
 LOCUS 601805465F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4056145 5',
 DEFINITION mRNA sequence.
 ACCESSION BF181843 GI:11059985
 VERSION BF181843.1
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Mus musculus; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 835)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
 http://image.llnl.gov
 Plate: L1AM9311 row: b column: 18
 High quality sequence stop: 672.
 Location/Qualifiers

FEATURES
 source
 1. 835
 /organism="Mus musculus"
 /strain="C57/B6"
 /db_xref="taxon:10090"
 /clone_image="4056145"
 /clone_lib="NCI_CGAP_Mam5"
 /tissue_type="tumor, gross tissue"
 /dev_stage="7 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-Sport6; Site:1; Salt:
 Site:2; Notti: Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by: Life Technologies. Investigator
 providing samples: Lothar Hennighausen/Robin Humphreys,
 NIH"

BASE COUNT 178 a 237 c 244 g 176 t
 ORIGIN

Query Match 26.9%; Score 401; DB 10; Length 835;
 Best Local Similarity 80.7%; Pred. NO. 1.3e-91;
 Matches 517; Conservative 0; Mismatches 120; Indels 4; Gaps 4;

QY 842 tcaactacatgacccagctgcagatgtgacacaaagatgaagacaggtgcaaaagt 901
 Db 8 TCAACTTCAAGACCAAGGTGCTCACTGTGACACCAAGGTGAGACATCAACAAAGAT 67
 QY 902 ccaagctgaacaggtacagctgcccagcaggtgctgacacaaagcgaagatcttg 961
 Db 68 CCAGGTGTACAGGTACCAAGTGTCCAGCAGAGGTCTGTGAGCAACAGGCGAAGGTC 126
 QY 962 gaagctctctatgaagatcgctcag -catatgcgcgcgcgcgcctcactcaggaac 1020
 Db 127 GCTCTCTTTTATGAAGTCTTCCAGCCATATGCCAGCTGCTATCCACTACGGTGC 186
 QY 1021 ctgagctgaaggaagcgtgtgatataccacgaagcgaaggaagtcctcctcttcgtg 1080
 Db 187 ATCGATGATCGAGGTGCTGCTGTGATCTCAACGAATGGAGATGATACCCCTCTTGTTC 246
 QY 1081 aagctgagagacagcgctgagctccctcagcaaatatacaacttcacgtcattcatg 1140
 Db 247 AAGTCTCAAAAAATGGCATGGATGCCCTGAGCAATACAGCCCTCTAGCTCTTCACT 306

QY 1141 gtgtcaaaagtgaaagtgacgatttgactgtctacacagacgtgtctcagctgccc 1200
 Db 307 GTGTCAAAAGTGACAGAG -ACGGCTGTGACTGCCACCGCACAGCTGTGCTCC 365
 QY 1201 ttgtgaagacgcaactcactcgtcccaagaatccaatgtgtccgacactgcaagacga 1260
 Db 366 TTGTGAGAGCCGGCAC -CACTGCCGAGAAATCCAGTGTCTGCGCTGTGTGAGAGAG 424
 QY 1261 ccttctactggctccgtgtgtttgaacccaacatctatgatactcactcaagcttgc 1320
 Db 425 CCACTCTATTGGGCTCTCTGTGTGAGACCAACATCTATGTGATCTCCAGCATTTGT 484
 QY 1321 aagagacgtgacgc 1380
 Db 485 AAGGCCCTGTGACGAGGTGTGATCTGAGAGAGGTGTGTGTGTGTGTGTGTGTGTGT 544
 QY 1381 cccgtgataaaagaaagacactcgtggtgctgctcagaatgagttcagtcgaagc 1440
 Db 545 CCGGTGTGCAAAAAGAGAGCTACGTGGCTCCCTCGAGAACGGGTGTGATCGAGAGC 604
 QY 1441 ctggagactcctcgagatgaaagccttcgcagatcttgc 1481
 Db 605 CTGAACTCTCTCAGAACGAAAGCCTTCGAGTCTTCG 645

RESULT 9
 AW910974 580 bp mRNA linear EST 25-MAY-2000-
 LOCUS ur81f08.y1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3156711 5'
 DEFINITION similar to TR:Q26422 Q26422 FACTOR C. [1] ;, mRNA sequence.
 ACCESSION AW910974
 VERSION AW910974.1 GI:8076219
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Mus musculus; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 580)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgp.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1992)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Jeffrey Green M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
 image.llnl.gov/image/html/resources.shtml

MG1:1059467
 Seq primer: -40RP from GIBCO
 High quality sequence stop: 428.
 Location/Qualifiers

FEATURES
 source
 1. 580
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone_image="3156711"
 /clone_lib="NCI_CGAP_Mam6"
 /sex="female; virgin"
 /tissue_type="infiltrating ductal carcinoma"
 /dev_stage="5 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-Sport6; Site:1; Salt:
 Site:2; Notti: Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by: Life Technologies. Investigator
 providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 138 a 150 c 163 g 129 t

DEFINITION IL5-MT0265-290301-421-a05 MT0265 Homo sapiens cDNA, mRNA sequence.
 ACCESSION B1030828
 VERSION B1030828.1 GI:14437458
 EST.
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 406)
 REFERENCE Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brustein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&ct=IL5-MT0265-290301-421-a05&f3=2001-03-29&f4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 354.
 Location/Qualifiers
 1. 406
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="MT0265"
 /note="Organ: marrow; Vector: puc18; Site:1; Sma1; Site:2; Sma1; A mini-library was made by cloning products derived from ORESPES PCR (U.S. Letters Patent application No. 196 from the Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 85 a 112 c 105 g 104 t
 ORIGIN
 Query Match 26.5%; Score 395; DB 10; Length 406;
 Best Local Similarity 99.8%; Pred. No. 3.1e-90;
 Matches 406; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1076 tctgtgaagcttgagagacacggtgtgcaagtcctcgaacaaataacacccctcagctcat 1135
 Db 406 TCGGAGAGCTGTGAGAGACAGGCGCTCAGCTCCAGCAAAATAC AACCTTCCAGCTCAT 348
 QY 1136 tcatgtgtcctaaagtgaaagctgacgatttgactgctacagacgctgtgtcagcgt 1195
 Db 347 TCAGTGTGTCTCAAAAGTAAAGTGCAGATTTGGACATGCTACAGACCGTGTGCTACACTGT 288
 QY 1196 gccgcttgaaagcagcaactcactgcccagaatccattgtccggacacatgcaag 1255
 Db 287 GCCCTTTGAAAGCCACACACTCAGTCCCAAGATTCATTGTCCGGCACACTGCAAG 228
 QY 1256 acgaacctctactgaggtcgcgtgtgttggaacacacatcatatgacgaatacctaaag 1315
 Db 227 ACGAACCTTCTACTAGGCTCCGCTGTGGACCAACATCATGCAAGATACCTCAAGCA 168
 QY 1316 tctgcaagacagctgtgcaagcgagatcatcagcaagagagtgaggtgtacgtggagc 1375
 Db 167 TCTGCAAGACAGCTGTGACAGCGCGGAGTCAATGACAGAGATGGGGGTGACGCGACG 108

QY 1376 tgaatccctgtgataaaagaagacactacgtggtgctcgcaggaatgagttcagttcg 1435
 Db 107 TCATGCTCCGTGTATATAAAGAGACCTACGTGGCTGCTCAGAAAGGATTCAGCTG 48
 QY 1436 aaagcttgagactcctcgtggtgagaaagcctcccgatcttgc 1482
 Db 47 AAAGCTGGGAGCTCTCGCATGGAAGGCTTCGGATCTTGGCT 1
 RESULT 12
 BFL14923 822 bp mRNA linear EST 24-OCT-2000
 LOCUS BFL190412F1 NCL_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4021274 5'
 DEFINITION mRNA sequence.
 ACCESSION BFL14923
 VERSION BFL14923.1 GI:10980963
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 822)
 REFERENCE NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA library preparation: Life Technologies, Inc.
 DNA sequencing by: Incycle Genomics, Inc.
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LML at:
 http://image.llnl.gov
 Plate: LLM9276 row: g column: 03
 High quality sequence stop: 666.
 Location/Qualifiers
 1. 822
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:4021274"
 /clone_lib="NCL_CGAP_Lu30"
 /clone_type="tumor, metastatic to mammary"
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: SalI; transgenic model MMTV-1, expression driven by MMTV-ltr enhancer; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies.
 Investigator providing samples: Gilbert Smith, NIH"
 BASE COUNT 196 a 209 c 250 g 167 t
 ORIGIN
 Query Match 26.2%; Score 391.2; DB 10; Length 822;
 Best Local Similarity 76.3%; Pred. No. 4.1e-89;
 Matches 575; Conservative 0; Mismatches 163; Indels 16; Gaps 7;
 QY 739 gaagagatgaatgagtgagaaagcgtccatctcctgaagaaacacatgttggctcaa 798
 Db 1 GATGTGATGAAGCGAGGAGGAAATCCGCCCTCGCCAGAGAAACCCATGCTGTGGTTCA 60
 QY 799 ccgaagtgatgagacccaagcccaagaacacctcgtcggtcaactacatgaccaa 858
 Db 61 CCCAGGCTGAT-----CAAGACCAAGAACCCCACTCATCAACCTTCAGACCCAA 111
 QY 859 gtcgtcagatgtgacacccaagatgaagagacagtgcaaggtgtccacgtgttaacagttac 918
 Db 112 GTGGTCCACTGTGACACCAAGATGAAGACATCAAGAAAGATCTCACGTTACAGTAC 171
 QY 919 cagtgcccaagcgtgctgctgaacacaaagcgaagatccttggaaatctgtctatgaa 978
 Db 111 cagtgcccaagcgtgctgctgacacacaaagcgaagatccttggaaatctgtctatgaa 978

AUTHORS
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, M. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.J.

TITLE
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL MEDLINE
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

COMMENT
 Contact: Simpson A.J.J.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV3<2=QV3-CT0556-041000-370-f05<3=2000-10-04<4=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 9
 High quality sequence stop: 410.
 Location/Qualifiers
 1. 410
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1lb="CT0556"
 /dev_stage="Adult"
 /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT
 109 a 111 c 109 g 81 t

ORIGIN

Query Match 25.7%; Score 383.6; DB 10; Length 410;
 Best Local Similarity 98.8%; Pred. No. 2.5e-87;
 Matches 397; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1005 catccactcggagatccctgatgacaaaggagcctgttgatatacccgagaaagaa 1064
 DB 9 CATCCACTCAGGAGATCCTGGATGACAAAGGAGCGCTGTGATATACACGAGGAGAA 68

QY 1065 ggtccctctctcgt-gaagctcgagagacagcgctgcagctccctcaagaataaacaac 1123
 DB 69 ggtccctctctcgtgagctcgagagacagcgctgcagctccctcaagaataaacaac 128

QY 1124 ctccagctcaatcatgtgtgtaaaagtgaaagtgcaagatttgagctgtactacagccg 1183
 DB 129 CTCCAGCTCATTCATGATGCTGTAAGTGAAGTGCAAGATTTCGACTGACAGACG 188

QY 1184 ttgtcgaagctgtcccgcttgtaaaagcgaactcactctgcccagaagatcatgttcgg 1243
 DB 189 TTGCTAGAGCTGTCCCGTTGAAAGCCAGCAACCTCACTGCCCCAAGATTCATTGTCGG 248

QY 1244 cacactgcaaaagagaactctcactctgagctcgggtgttgtaacaacaatcatatgag 1303
 DB 249 CACACTGCAAAAGAGAACTTCTCTACTGGGCTCCGGTGTGGAAACAACATCTATGAG 308

QY 1304 ataactcaagcatctcgaagacagctgtgcacgcgggaagtcatacgaagagagtgagg 1363
 DB 309 ATAACCTCAAGCATCTGCAAGACAGCTGTGACGAGAGATCATCAGCAACGAGATGGGG 368

QY 1364 gtgaagctgacgtgatgacccgttgataaaagaaagacttaccg 1405
 DB 369 GTGACGTGAGACGTGATGCCCGGTGGATCAAAAGAAAGACCTTACG 410

RESULT 15
 BG661553

LOCUS
 DEFINITION BG661553 763 bp mRNA linear EST 29-MAY-2001
 602795324P2 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4916597 5', mRNA sequence.

ACCESSION
 BG661553

VERSION
 BG661553.1 GI:14212091

KEYWORDS
 EST.

SOURCE
 house mouse.

ORGANISM
 Mus musculus

REFERENCE
 1 (bases 1 to 763)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1995).

AUTHORS
 JOURNAL

COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bs-r@mail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at:
<http://image.llnl.gov>
 Plate: LIML0826 row: p column: 06
 High quality sequence stop: 755.
 Location/Qualifiers
 1. 763
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_1lb="IMAGE:4916597"
 /clone_1lb="NIH_CGAP_Mam4"
 /tissue_type="tumor, gross tissue"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: PCMV-SPORE6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: oligo dT. Average insert 2.5 kb. Library constructed by Life Technologies, catalog # 12018-016. Investigators providing samples: Lothar Hennighausen/Priscilla Furth, NIH Reference for transgenic model: Li et al., Cell Growth and Differentiation 7, 3-11 (1996). Note: this is a NIH_CGAP Library."

BASE COUNT
 165 a 238 c 213 g 147 t

ORIGIN

Query Match 25.4%; Score 379.2; DB 10; Length 763;
 Best Local Similarity 78.1%; Pred. No. 4.5e-86;
 Matches 481; Conservative 0; Mismatches 133; Indels 2; Gaps 2;

QY 1 atgagctgcgtccctgggtgtgtcatcccttggggctgtctctcgtgtcgtatcc 60
 DB 149 ATGAGCTGCTTCTGTAACATATGCTGATGGGGCTGCTGTGCTGTGTGAGTA 208

QY 61 caagctactcctctgccaagctcactctcttagagagagctgtctcagcaataaccagcac 120
 DB 209 CAGGCTTTTTCCTTCCCAACACACAGAGCTGTGAGAGTGTCTGAGCAATATACCACTAT 268

QY 121 aacgaatcactcccggtgccgaagacatcccccagagagaggaagagagatccctc 180
 DB 269 GCAGAGCCACACACTCAGCGGTGGCGAGGCGCATGCCATGTCAAGACGCCCAAGATCTCTC 328

QY 181 atgtgtcaacaagaagctctcgggccaggtgtgcagctcctcagcctcaacatgtgacatg 240
 DB 329 ATGCTCACACACAGAGCTGGAGGCGCAGGTGTATCCCTCTCCCAACATGGAACACATG 388

QY 241 acctggatgagaaacttgagagagctgtgctgcagcgtgtggccagctgcatctatggag 300
 DB 389 ACTTGGATGAGAGCTAGAGAGATCTGCAGCAGCGTGGCGACATATAGATGCTGTGGAG 448

Tue May 28 16:10:01 2002

us-09-667-380a-1.ln1

No 34 contg. 859
Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 25, 2002, 10:33:09 ; Search time 70.13 Seconds
(without alignments)
5222.295 Million cell updates/sec

Title: US-09-667-380A-1
1491

Sequence: 1 atgagctgcgcctcctggtggtg.....ggatcttctcgtcagcag 1491

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/prodata/2/lna/5A.COMB.seq:*
2: /cgn2_6/prodata/2/lna/5B.COMB.seq:*
3: /cgn2_6/prodata/2/lna/6A.COMB.seq:*
4: /cgn2_6/prodata/2/lna/6B.COMB.seq:*
5: /cgn2_6/prodata/2/lna/PCUS.COMB.seq:*
6: /cgn2_6/prodata/2/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	74.6	5.0	970	2	US-08-773-368-2
2	74.6	5.0	970	3	US-09-199-887-2
3	43.2	2.9	7218	1	US-08-232-463-14
4	42.2	2.8	11219	3	US-07-642-734C-1
5	42.2	2.8	11219	3	US-08-439-009A-1
6	38.4	2.6	696	4	US-08-998-416-689
7	38.4	2.6	30001	2	US-08-125-468-1
8	38.4	2.6	30001	2	US-08-474-933-1
9	37.4	2.5	4031	1	US-08-159-784-1
10	37	2.5	1211	2	US-08-997-080-40
11	37	2.5	1211	2	US-08-997-362-40
12	37	2.5	1211	2	US-08-873-970-40
13	37	2.5	1211	4	US-09-095-855-40
14	37	2.5	1211	4	US-08-705-347A-40
15	37	2.5	1211	4	US-09-324-542-40
16	36.8	2.5	756	4	US-09-175-014-1
17	36.4	2.4	3468	1	US-07-951-715A-2
18	36.4	2.4	3468	2	US-08-459-448A-2
19	36.4	2.4	3468	3	US-08-459-595A-2
20	36.4	2.4	3468	3	US-08-459-504B-2
21	36.4	2.4	3468	3	US-08-459-444-2
22	36.4	2.4	3468	3	US-09-053-549-3
23	36.4	2.4	3468	4	US-09-547-422-2
24	36	2.4	1212	6	5212296-17
25	35.8	2.4	1410	3	PCT-US95-15428-14
26	35.8	2.4	1817	3	US-08-564-496C-14
27	35.6	2.4	696	4	US-08-998-416-1050

28	35.2	2.4	71989	4	US-09-443-501A-2	Sequence 2, Appl
29	34.8	2.3	1080	2	US-09-205-922-1	Sequence 1, Appl
30	34.8	2.3	1947	1	US-07-951-715A-3	Sequence 3, Appl
31	34.8	2.3	1947	2	US-08-459-448A-3	Sequence 3, Appl
32	34.8	2.3	1947	3	US-08-459-595A-3	Sequence 3, Appl
33	34.8	2.3	1947	3	US-08-459-504B-3	Sequence 3, Appl
34	34.8	2.3	1947	4	US-09-547-422-3	Sequence 3, Appl
35	34.8	2.3	1947	4	US-07-951-715A-4	Sequence 4, Appl
36	34.8	2.3	3468	1	US-08-459-448A-4	Sequence 4, Appl
37	34.8	2.3	3468	2	US-08-459-595A-4	Sequence 4, Appl
38	34.8	2.3	3468	2	US-08-459-504B-4	Sequence 4, Appl
39	34.8	2.3	3468	3	US-08-459-444-4	Sequence 4, Appl
40	34.8	2.3	3468	3	US-08-459-504B-4	Sequence 4, Appl
41	34.8	2.3	3468	3	US-08-459-444-4	Sequence 4, Appl
42	34.8	2.3	3468	3	US-08-459-504B-4	Sequence 4, Appl
43	34.8	2.3	3468	3	US-08-459-444-4	Sequence 4, Appl
44	34.8	2.3	3468	3	US-08-459-504B-4	Sequence 4, Appl
45	34.8	2.3	3468	3	US-08-459-444-4	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-08-773-368-2
; Sequence 2, Application US/08773368
; Patent No. 5856130
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goll, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN PATHOGENESIS-RELATED
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Inocyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/773,368
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy RJ
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PP-0186 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 970 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: 1599164
; US-08-773-368-2
Query Match 5.0%; Score 74.6; DB 2; Length 970;
Best local Similarity 51.2%; Pred. No. 3.9e-11;

Matches 275; Conservative 0; Mismatches 226; Indels 36; Gaps 3;

```

QY 146 gagccatcccccagggaggaagagagatcctcatgtgcacaacaagcttcgggccc 205
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 88 GAGCCCTCACAGATGAGGAGAAACGTTTGATGTGAGCTGCACAAACCTTACCGGGCCC 147
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 206 aggtgagcctcagcagcctcccaaatagatgacatgagctggagctgagctgagag 265
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 148 AGGTATCCCGAGGAGGCTCAGACATGCTGACATGAGATGGAGAGGAGAGCTGCGCCCT 207
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 266 ctgtcagcagctgggccaatgcatgcatctggagcagcagcccaagctcgtgtgt 325
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 208 TCGCCAGGCTTACGACGAGGAGTTCGTNG-----GGCCACAAACAGAGAGCGCGGC 261
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 326 ccatgggagagaaactggcgctcactgaggagagatgctcctccgggttcacatgtc 385
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 262 GCCCGGCGAGAAATCTGTTCGCCATCACAGAGGAGATGAGCTGCCGCTGGCCATGG 321
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 386 agtctgtatgacgaggtgagagctacactaccctaccgagcagatgcaacct 445
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 322 AGGAGTGCCACACGAGCTGACACCTAACCTCAGCGCGCCACCT----- 369
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 446 ggttcacagagagctgctgggacctatgtgacgacactacacacagatgttggcca 505
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 370 -----GCAGCCAGGCGCAGATGTGCGCCACTACACGAGGTGATGGCCA 417
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 506 ccacacaagaatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 565
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 418 AGACAGAGAGATGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 471
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 566 ttgggagagagcgggtacttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 625
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 472 AGGAGACCAACATCGAATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 531
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 626 aagccccctacaagaatgcccgcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 682
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 532 AACGCCCTTACGACGAGGAGACTCCGTGCTCCCAATGTCTGCTGCTACACTGCA 588
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 2

US-09-199-887-2

Sequence 2, Application US/09199887

Patent No. 6071874

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

APPLICANT: GOLI, Surya K.

TITLE OF INVENTION: NOVEL HUMAN PATHOGENESIS-RELATED

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESS: Indyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/199,887

CLASSIFICATION:

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 08/773,368

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy RJ

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0186 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 970 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

CLONE: 1599164

US-09-199-887-2

Query Match

Best Local Similarity 5.0%; Score 74.6; DB 3; Length 970;

Matches 275; Conservative 0; Mismatches 226; Indels 36; Gaps 3;

```

QY 146 gagccatcccccagggaggaagagagatcctcatgtgcacaacaagcttcgggccc 205
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 88 GAGCCCTCACAGATGAGGAGAAACGTTTGATGTGAGCTGCACAAACCTTACCGGGCCC 147
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 206 aggtgagcctcagcagcctcccaaatagatgacatgagctggagctgagctgagag 265
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 148 AGGTATCCCGAGGAGGCTCAGACATGCTGACATGAGATGGAGAGGAGAGCTGCGCCCT 207
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 266 ctgtcagcagctgggccaatgcatgcatctggagcagcagcccaagctcgtgtgt 325
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 208 TCGCCAGGCTTACGACGAGGAGTTCGTNG-----GGCCACAAACAGAGAGCGCGGC 261
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 326 ccatgggagagaaactggcgctcactgaggagagatgctcctccgggttcacatgtc 385
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 262 GCCCGGCGAGAAATCTGTTCGCCATCACAGAGGAGATGAGCTGCCGCTGGCCATGG 321
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 386 agtctgtatgacgaggtgagagctacactaccctaccgagcagatgcaacct 445
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 322 AGGAGTGCCACACGAGCTGACACCTAACCTCAGCGCGCCACCT----- 369
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 446 ggttcacagagagctgctgggacctatgtgacgacactacacacagatgttggcca 505
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 370 -----GCAGCCAGGCGCAGATGTGCGCCACTACACGAGGTGATGGCCA 417
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 506 ccacacaagaatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 565
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 418 AGACAGAGAGATGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 471
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 566 ttgggagagagcgggtacttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 625
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 472 AGGAGACCAACATCGAATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 531
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 626 aagccccctacaagaatgcccgcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 682
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 532 AACGCCCTTACGACGAGGAGACTCCGTGCTCCCAATGTCTGCTGCTACACTGCA 588
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 3

US-08-232-463-14/C

Sequence 14, Application US/08232463

Patent No. 5670367

GENERAL INFORMATION:

APPLICANT: DORNER, F.

APPLICANT: SCHRIFFLINGER, F.

TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESS: Foley & Lardner

STREET: 1800 diagonal Road, Suite 500

CITY: Alexandria

STATE: VA

COUNTRY: USA

ZIP: 22313-0299

COMPUTER READABLE FORM:

GENERAL INFORMATION:
APPLICANT: Katz, L
APPLICANT: Donadio, S
APPLICANT: Mcalpine, J B
TITLE OF INVENTION: Recombinant DNA Method for Producing
TITLE OF INVENTION: Pyrimycin Analogs
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSER: Edward H. Gorman
STREET: Abbott Laboratories D377/AP6D-2 One Abbott
STREET: Park Rd
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/642.734C
FILING DATE: 17-JAN-91
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Danckers, Andreas M
REGISTRATION NUMBER: 32652
REFERENCE/DOCKET NUMBER: 4952.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9196
TELEFAX: 708-938-4623
INFORMATION FOR SEQ ID NO: 1:
INFORMATION CHARACTERISTICS:

```

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 2338
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 744..6659
  OTHER INFORMATION: /function= "APPROXIMATE SPAN OF
  OTHER INFORMATION: MODULE 1"
  OTHER INFORMATION: /label= FUNCTION
FEATURE:
  NAME/KEY: CDS
  LOCATION: 744..11219
  OTHER INFORMATION: /function= "gene= "eryA"
  OTHER INFORMATION: /product= "ORF1 encoding modules 1 & 2 for
  OTHER INFORMATION: 6'-deoxyerythronolide B""
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 744..1868
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: acyltransferase domain 1 of module 1"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 1998..2198
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: acyl carrier domain 1 of module 1"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 2250..3626
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: beta-ketoacylACP synthase domain/module1"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 3831..4811

```

ADDRESSSEE: Steven P. Weinstein
STREET: Abbott Laboratories D377/AP6D-2 One Abbott
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439,009A
FILING DATE: 11-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Casuto, Dianne
REGISTRATION NUMBER: 40,943
REFERENCE/DOCKET NUMBER: 4952.US.D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847-938-3137
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 11219 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 2338
FEATURE:

```

Query March
Best Local Similarity 2.88; Score 42.2; DB 1; Length 11219;
Matches 98; Conservative 51.3%; Pred. No. 0.076; Mismatches 93; Indels 0; Gaps 0

Oy 199 cgggagcaagtgctcagcgtccagcctcccaatagagatgaatgaactggatgacgaactg 258
Db 8324 CGGGAAACAATCCCGCGCCATGCCTTCACACTGGACACCTGGCCAGGAAAGCAAAAC 258
Oy 259 gagaagctctgctcagagtgtyggcgcaatcagtcgatactggagacagcgccacagctcg 318
Db 8264 TGAACCAACGCTGGAGAGCACTGCCGGGTGCAACACCGGGAAACACCGCGCTCGGCCAC 8205
Oy 319 ctggtgtcatctggggaagaacctggagcctgacgtgacactggggcaggtatcgtctccggatlc 378
Db 8204 CGCCGCCAACCGGTGACGCGCTCGCGGTCGCGCGGAGAACACCGCGCGGCGGCTCTC 8145
Oy 379 catgtgcagtc 389
Db 8144 GAAGTCCTGGC 8134

RESULT 5
US-08-439-009A-1/c
; Sequence 1, Application US/08439009A
; Patent No. 6004787
; GENERAL INFORMATION:
; APPLICANT: Donadio, S
; APPLICANT: Katz, L
; APPLICANT: Mcalpine, J B
; TITLE OF INVENTION: Method of Directing Biosynthesis of
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:

```

```

1 OTHER INFORMATION: /function= "APPROXIMATE SPAN OF
2 OTHER INFORMATION: MODULE 1"
3 OTHER INFORMATION:
4 FEATURE:
5 NAME/KEY: CDS
6 LOCATION: 744..11219
7 OTHER INFORMATION: /function= "gene= "eryA""
8 OTHER INFORMATION: /product= "GRF1 encoding modules 1 & 2 for
9 OTHER INFORMATION: 6-deoxyerythronolide B"
10 FEATURE:
11 NAME/KEY: misc_feature
12 LOCATION: 744..1868
13 OTHER INFORMATION: /function= "approximate span of
14 OTHER INFORMATION: acyltransferase domain 1 of module 1"
15 FEATURE:
16 NAME/KEY: misc_feature
17 LOCATION: 1998..2198
18 OTHER INFORMATION: /function= "approximate span of
19 OTHER INFORMATION: acyl carrier domain 1 of module 1"
20 FEATURE:
21 NAME/KEY: misc_feature
22 LOCATION: 2250..3626
23 OTHER INFORMATION: /function= "approximate span of
24 OTHER INFORMATION: beta-ketocacylACP synthase domain/module1"
25 FEATURE:
26 NAME/KEY: misc_feature
27 LOCATION: 3831..4811
28 OTHER INFORMATION: /function= "approximate span of
29 OTHER INFORMATION: acyltransferase 2 domain of module 1"
30 FEATURE:
31 NAME/KEY: misc_feature
32 LOCATION: 5574..6125
33 OTHER INFORMATION: /function= "approximate span of
34 FEATURE: beta-ketoreductase domain of module 1"
35 NAME/KEY: misc_feature

```

```

1: ZIP: 27709
2: COMPUTER READABLE FORM:
3:
4: MEDIUM TYPE: Floppy disk
5: COMPUTER: IBM PC compatible
6: OPERATING SYSTEM: PC-DOS/MS-DOS
7: SOFTWARE: Patentin Release #1.0, Version #1.30
8:
9: CURRENT APPLICATION DATA:
10: APPLICATION NUMBER: US/08/998,416
11: FILING DATE: 24-DEC-1997
12: CLASSIFICATION: 435
13:
14: PRIORITY APPLICATION DATA:
15: PRIOR APPLICATION NUMBER: CH 0016/97
16: FILING DATE: 31-DEC-1996
17: ATTORNEY/AGENT INFORMATION:
18: NAME: Meigs, J. Timothy
19: REGISTRATION NUMBER: 38,241
20: REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
21: TELECOMMUNICATION INFORMATION:
22: TELEPHONE: 919-541-8587
23: TELEFAX: 919-541-8689
24:
25: INFORMATION FOR SEO ID NO: 689:
26: SEQUENCE CHARACTERISTICS:
27: LENGTH: 696 base pairs
28: TYPE: nucleic acid
29: STRANDEDNESS: single
30: TOPOLOGY: linear
31: MOLECULE TYPE: DNA (genomic)
32: ORIGINAL SOURCE:
33: ORGANISM: PAG1456RP
34:
35: US-08-998-416-689
36:
37:
38:
39:
40:
41:
42:
43:
44:
45:
46:
47:
48:
49:
50:
51:
52:
53:
54:
55:
56:
57:
58:
59:
60:
61:
62:
63:
64:
65:
66:
67:
68:
69:
70:
71:
72:
73:
74:
75:
76:
77:
78:
79:
80:
81:
82:
83:
84:
85:
86:
87:
88:
89:
90:
91:
92:
93:
94:
95:
96:
97:
98:
99:
100:
101:
102:
103:
104:
105:
106:
107:
108:
109:
110:
111:
112:
113:
114:
115:
116:
117:
118:
119:
120:
121:
122:
123:
124:
125:
126:
127:
128:
129:
130:
131:
132:
133:
134:
135:
136:
137:
138:
139:
140:
141:
142:
143:
144:
145:
146:
147:
148:
149:
150:
151:
152:
153:
154:
155:
156:
157:
158:
159:
160:
161:
162:
163:
164:
165:
166:
167:
168:
169:
170:
171:
172:
173:
174:
175:
176:
177:
178:
179:
180:
181:
182:
183:
184:
185:
186:
187:
188:
189:
190:
191:
192:
193:
194:
195:
196:
197:
198:
199:
200:
201:
202:
203:
204:
205:
206:
207:
208:
209:
210:
211:
212:
213:
214:
215:
216:
217:
218:
219:
220:
221:
222:
223:
224:
225:
226:
227:
228:
229:
230:
231:
232:
233:
234:
235:
236:
237:
238:
239:
240:
241:
242:
243:
244:
245:
246:
247:
248:
249:
250:
251:
252:
253:
254:
255:
256:
257:
258:
259:
260:
261:
262:
263:
264:
265:
266:
267:
268:
269:
270:
271:
272:
273:
274:
275:
276:
277:
278:
279:
280:
281:
282:
283:
284:
285:
286:
287:
288:
289:
290:
291:
292:
293:
294:
295:
296:
297:
298:
299:
300:
301:
302:
303:
304:
305:
306:
307:
308:
309:
310:
311:
312:
313:
314:
315:
316:
317:
318:
319:
320:
321:
322:
323:
324:
325:
326:
327:
328:
329:
330:
331:
332:
333:
334:
335:
336:
337:
338:
339:
340:
341:
342:
343:
344:
345:
346:
347:
348:
349:
350:
351:
352:
353:
354:
355:
356:
357:
358:
359:
360:
361:
362:
363:
364:
365:
366:
367:
368:
369:
370:
371:
372:
373:
374:
375:
376:
377:
378:
379:
380:
381:
382:
383:
384:
385:
386:
387:
388:
389:
390:
391:
392:
393:
394:
395:
396:
397:
398:
399:
400:
401:
402:
403:
404:
405:
406:
407:
408:
409:
410:
411:
412:
413:
414:
415:
416:
417:
418:
419:
420:
421:
422:
423:
424:
425:
426:
427:
428:
429:
430:
431:
432:
433:
434:
435:
436:
437:
438:
439:
440:
441:
442:
443:
444:
445:
446:
447:
448:
449:
450:
451:
452:
453:
454:
455:
456:
457:
458:
459:
460:
461:
462:
463:
464:
465:
466:
467:
468:
469:
470:
471:
472:
473:
474:
475:
476:
477:
478:
479:
480:
481:
482:
483:
484:
485:
486:
487:
488:
489:
490:
491:
492:
493:
494:
495:
496:
497:
498:
499:
500:
501:
502:
503:
504:
505:
506:
507:
508:
509:
510:
511:
512:
513:
514:
515:
516:
517:
518:
519:
520:
521:
522:
523:
524:
525:
526:
527:
528:
529:
530:
531:
532:
533:
534:
535:
536:
537:
538:
539:
540:
541:
542:
543:
544:
545:
546:
547:
548:
549:
550:
551:
552:
553:
554:
555:
556:
557:
558:
559:
560:
561:
562:
563:
564:
565:
566:
567:
568:
569:
570:
571:
572:
573:
574:
575:
576:
577:
578:
579:
580:
581:
582:
583:
584:
585:
586:
587:
588:
589:
590:
591:
592:
593:
594:
595:
596:
597:
598:
599:
600:
601:
602:
603:
604:
605:
606:
607:
608:
609:
610:
611:
612:
613:
614:
615:
616:
617:
618:
619:
620:
621:
622:
623:
624:
625:
626:
627:
628:
629:
630:
631:
632:
633:
634:
635:
636:
637:
638:
639:
640:
641:
642:
643:
644:
645:
646:
647:
648:
649:
650:
651:
652:
653:
654:
655:
656:
657:
658:
659:
660:
661:
662:
663:
664:
665:
666:
667:
668:
669:
670:
671:
672:
673:
674:
675:
676:
677:
678:
679:
680:
681:
682:
683:
684:
685:
686:
687:
688:
689:
690:
691:
692:
693:
694:
695:
696:
697:
698:
699:
700:
701:
702:
703:
704:
705:
706:
707:
708:
709:
710:
711:
712:
713:
714:
715:
716:
717:
718:
719:
720:
721:
722:
723:
724:
725:
726:
727:
728:
729:
730:
731:
732:
733:
734:
735:
736:
737:
738:
739:
740:
741:
742:
743:
744:
745:
746:
747:
748:
749:
750:
751:
752:
753:
754:
755:
756:
757:
758:
759:
760:
761:
762:
763:
764:
765:
766:
767:
768:
769:
770:
771:
772:
773:
774:
775:
776:
777:
778:
779:
780:
781:
782:

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/125,468
FILING DATE: 22-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Tsevdos, Estelle J
REGISTRATION NUMBER: 31,145
REFERENCE/DOCKET NUMBER: 31,255-02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201)831-3241
TELEFAX: (201)831-3305
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 30001 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-125-468-1

Query Match
Best Local Similarity 2.6%; Score 38.4; DB 1; Length 30001;
Matches 87; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

DB 160 gaggaagaaggaagatcctcatgtgtcacacaagaagcttcggggccaggtgacagcctcag 219
24305 GTGGCCGGCGAGGTGAACCTTCACCGCGAGGACACCTTCGCGCGCGACGTCTGCCGACG 24246

QY 220 gctccacaatggaagatgacatgacgtggatgacgaactggaagagtgctgtgcaagctg 279
24245 ACCGACACATGACGCGCGCTGCTGTGCGCGCGAGGAGGCGCTCGCGCGAGCGGGG 24186

QY 280 gccagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 327
24185 GCGGACCGACGACGAGTCCGACTTCTCCGCCGCGCTGTACAGGCC 24138

RESULT 8
US-08-474-933-1/C
Sequence 1, Application US/08474933
Patent No. 5866410
GENERAL INFORMATION:
APPLICANT: Ryan, Michael J.
APPLICANT: Lotvin, Jason A.
APPLICANT: Strathy, Nancy E.
APPLICANT: Fantini, Susan E.
TITLE OF INVENTION: Cloning of the biosynthetic pathway for
TITLE OF INVENTION: chlorotetracycline and tetracycline formation and cosmid
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: American Cyanamid Company
STREET: One Cyanamid Plaza
CITY: Wayne
STATE: New Jersey
COUNTRY: USA
ZIP: 07470
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,933
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/125,468
FILING DATE: 22-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Tsevdos, Estelle J
REGISTRATION NUMBER: 31,145
REFERENCE/DOCKET NUMBER: 31,255-02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201)831-3241
TELEFAX: (201)831-3305
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 30001 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-474-933-1

Query Match
Best Local Similarity 2.6%; Score 38.4; DB 2; Length 30001;
Matches 87; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

DB 160 gaggaagaaggaagatcctcatgtgtcacacaagaagcttcggggccaggtgacagcctcag 219
24305 GTGGCCGGCGAGGTGAACCTTCACCGCGAGGACACCTTCGCGCGCGACGTCTGCCGACG 24246

QY 220 gctccacaatggaagatgacatgacgtggatgacgaactggaagagtgctgtgcaagctg 279
24245 ACCGACACATGACGCGCGCTGCTGTGCGCGCGAGGAGGCGCTCGCGCGAGCGGGG 24186

QY 280 gccagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 327
24185 GCGGACCGACGACGAGTCCGACTTCTCCGCCGCGCTGTACAGGCC 24138

RESULT 9
US-08-159-784-1/C
Sequence 1, Application US/08159784
Patent No. 5643783
GENERAL INFORMATION:
APPLICANT: Bjorn R. Olsen
TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: MordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/159,784
FILING DATE: December 1, 1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: John F. Freeman
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00246/170001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 4031
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-159-784-1

Query Match 2.5%; Score 37.4; DB 1; Length 4031;
Best Local Similarity 52.2%; Pred. No. 1;
Matches 83; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 402 ggtgaagagctacacacccctaccgagcgagtgcaacccctggtgtccagagaagt 461
DB 1087 ggtgaagagctacacacccctaccgagcgagtgcaacccctggtgtccagagaagt 461
QY 462 ctgagagctatgtgacgacacacagatagtttgagccacacacacagatcg 521
DB 1027 ctgagagctatgtgacgacacacagatagtttgagccacacacacagatcg 521
QY 522 ttgtctgtgacacacctgctcggaagatgactgtctggg 560
DB 967 cagagacctggggccagctgctggccagctggccctggg 929

RESULT 10
US-08-997-080-40
Sequence 40, Application US/08997080
Patent No. 5968524

GENERAL INFORMATION:

APPLICANT: WATSON, JAMES D.
APPLICANT: TAN, PAUL L.J.
TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
NUMBER OF SEQUENCES: 194
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/997,080

FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:

NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1007
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563

TELEX:

INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 1211 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-997-080-40

Query Match 2.5%; Score 37; DB 2; Length 1211;
Best Local Similarity 44.6%; Pred. No. 0.84; Indels 0; Gaps 0;
Matches 145; Conservative 0; Mismatches 180;

QY 119 acaagagctcactcccggtctcgcagagcatcccccaggaaggaagagatcc 178
DB 383 ACACCGACTGGTAGCGCCCGCCGTAACAAGAGCCCGACCGTAGACTCAAGTAGGAGA 442
QY 179 tcatgtcacacaagcttcgggccaagtgacagctcagacctcaaatgagta 238
DB 443 CCTTCTGACCCAGAGAGCTCCCGGCTGCTGACAGGCCCAACCCGCGTCAAGCCAGCC 502
QY 239 tgacctggaagaagcttgagaagcttgctgcagcgctgggccaagtcagtcagtc 298
DB 503 GCAGCGCCCTTCGCTGCTGATGAGCGGCTTCGCGCCGCTGACCTGAGACTGCG 562
QY 299 agcagcgcccaacacactgctgctgctcagctggcagaaactgggctcagtcag 358
DB 563 ACCCGAGAGCTCATCTACGCGGCTCGATGCTCGGCTTCTGAACTCCAGAGGCT 622
QY 359 ggtatgcttcggggttcacatgtgcagctctggtatgacagagtgaaagtaacct 418
DB 623 GGTGGCGCTTCGATCAACATCTCGATGAGCGCGCGGCTTCAAGCGCGAGCA 682
QY 419 accctacccgagcgagtgcaacc 443
DB 683 TGTGGCGCAAGACCGAGGATCCC 707

RESULT 11
US-08-997-362-40
Sequence 40, Application US/08997362
Patent No. 5985287

GENERAL INFORMATION:

APPLICANT: Tan, Paul
APPLICANT: Miyama, Jun
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Scott, Linda
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
NUMBER OF SEQUENCES: 194
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/997,362

FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
FILING DATE: June 12, 1997
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347

ATTORNEY/AGENT INFORMATION:

NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002c2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-263-0563

TELEX:

INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 1211 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-997-362-40

Query Match
Best Local Similarity 2.5%; Score 37; DB 2; Length 1211;
Matches 145; Conservative 0; Mismatches 180; Indels 0; Gaps 0;
QY 119 acaagagtcaccccggtcccgagagccatcccccagggagagagagatcc 178
DB 383 ACACCCAGCTGTACGCCGCCCGCCTAACAGAGGCCCGCCTGACCTACAACTGGGAGA 442
QY 179 tcatgtcacacaagagcttcgggcccagggtgcaagcctcagagcctcaactgtgagta 238
DB 443 CCTCTGTACCCAGAGAGCTCCCGGCTGGCTGACAGCCACCGCGGCTCAAGCCGACCG 502
QY 239 tgacctggatgacgaactggagagctgtgtcagagctggagcagtcagtcagtcag 298
DB 503 GCAGCGGCCCTGTGCTGTGATGGGGGCTTGGCGCTGACCTGGCGACCTGGC 562
QY 299 agcagggccccacagctgtgtgtccatcgagcagaaactggggcgctcactgggca 358
DB 563 ACCGGAGCACTTATCTACGCGGGCTCGATGTCGGCTTCTGAAACCCCTCCGAGGGCT 622
QY 359 ggtatgctctcgggggtccatgtgcaagctcgtgataagcaggtgaagactaacct 418
DB 623 GGTGGCGCTTCTGATCAACATCTCGATGGGTGACCGCGGCTTCAAGCGCGACGACA 682
QY 419 accctaccggagagtgaaacc 443
DB 683 TGTGGGGCAAGACGAGGGGATCCC 707

RESULT 12

US-08-873-970-40
Sequence 40, Application US/08873970
Patent No. 6001361

GENERAL INFORMATION:

APPLICANT: Tan, Paul
APPLICANT: Hiyma, Jun
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/873,970
FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/705,347
FILING DATE: 29-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sleeth, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002C1
TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563

INFORMATION FOR SEQ ID NO: 40:

SEQUENCE CHARACTERISTICS:
LENGTH: 1211 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-873-970-40

Query Match
Best Local Similarity 2.5%; Score 37; DB 3; Length 1211;
Matches 145; Conservative 0; Mismatches 180; Indels 0; Gaps 0;
QY 119 acaagagtcaccccggtcccgagagccatcccccagggagagagatcc 178
DB 383 ACACCCAGCTGTACGCCGCCCGCCTAACAGAGGCCCGCCTGACCTACAACTGGGAGA 442
QY 179 tcatgtcacacaagagcttcgggcccagggtgcaagcctcagagcctcaactgtgagta 238
DB 443 CCTCTGTACCCAGAGAGCTCCCGGCTGGCTGACAGCCACCGCGGCTCAAGCCGACCG 502
QY 239 tgacctggatgacgaactggagagctgtgtcagagctggagcagtcagtcagtcag 298
DB 503 GCAGCGGCCCTGTGCTGTGATGGGGGCTTGGCGCTGACCTGGCGACCTGGC 562
QY 299 agcagggccccacagctgtgtgtccatcgagcagaaactggggcgctcactgggca 358
DB 563 ACCGGAGCACTTATCTACGCGGGCTCGATGTCGGCTTCTGAAACCCCTCCGAGGGCT 622
QY 359 ggtatgctctcgggggtccatgtgcaagctcgtgataagcaggtgaagactaacct 418
DB 623 GGTGGCGCTTCTGATCAACATCTCGATGGGTGACCGCGGCTTCAAGCGCGACGACA 682
QY 419 accctaccggagagtgaaacc 443
DB 683 TGTGGGGCAAGACGAGGGGATCCC 707

RESULT 13

US-09-095-855-40
Sequence 40, Application US/09095855
Patent No. 6160093

GENERAL INFORMATION:

APPLICANT: Tan, Paul
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: Compounds and Methods for
NUMBER OF SEQUENCES: 208
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/095,855
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/705,347
FILING DATE: 29-AUG-1996

Tue May 28 16:10:01 2002

us-09-667-380a-1.rni

Page 9

APPLICATION NUMBER: 08/873,970
FILING DATE: 12-JUN-1997
APPLICATION NUMBER: 08/997,362
FILING DATE: 23-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002c3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 1211 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-095-855-40

Query Match 2.5%; Score 37; DB 4; Length 1211;
Best Local Similarity 44.6%; Pred. No. 0.84; Indels 0; Gaps 0;
Matches 145; Conservative 0; Mismatches 180;

OY 119 acaagaggtctactccgggtccgcagagcattcccccagagagagagagagatcc 178
DB 383 ACACCGACTGTACGCCGCCGCCCTTAACAAGGCGCCGACCGTACAAAGTGGAGA 442
OY 179 tcatgtcgaacaagaattcggggccaggtgagagctcagctcagctcgaacatgagata 238
DB 443 CTTCTGACCCAGGAGCTCCGGGCTGCGAGGCGCAACCGCGCGCTCAAGCCGACG 502
OY 239 tgacttgatgag 298
DB 503 GCAGGCGCCCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 562
OY 299 agcagggcccaacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 358
DB 563 ACCCGGAGAGTTCATTCACGCGGCTCGATGCCGCTTCTTGAACCCCTCCGAGGGCT 622
OY 359 ggtatgcctcgcgggttcctcatgtgcaagtcctggtatgagagagagagagag 418
DB 623 GGTGGCCGTTCTCTGATCAACATCTGATGAGGAGCGCGCGCTTCAAGCGCAGACA 682
OY 419 accctaccag 443
DB 683 TGTGGGCAAGACGAGGAGGATCCC 707

RESULT 14
US-08-705-347A-40
Sequence 40, Application US/08705347A
Patent No. 6284255
GENERAL INFORMATION:
APPLICANT: Tan, Paul
APPLICANT: Hiyma, Jun
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND
TITLE OF INVENTION: DIAGNOSIS OF MICROBACTERIAL INFECTIONS
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Spectran Picard PLLC
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/705,347A
FILING DATE: 28-AUG-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206,269,0565
TELEFAX: 206,269,0563
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 1211 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-705-347A-40

Query Match 2.5%; Score 37; DB 4; Length 1211;
Best Local Similarity 44.6%; Pred. No. 0.84; Indels 0; Gaps 0;
Matches 145; Conservative 0; Mismatches 180;

OY 119 acaagaggtctactccgggtccgcagagcattcccccagagagagagagagatcc 178
DB 383 ACACCGACTGTACGCCGCCGCCCTTAACAAGGCGCCGACCGTACAAAGTGGAGA 442
OY 179 tcatgtcgaacaagaattcggggccaggtgagagctcagctcagctcgaacatgagata 238
DB 443 CTTCTGACCCAGGAGCTCCGGGCTGCGAGGCGCAACCGCGCGCTCAAGCCGACG 502
OY 239 tgacttgatgag 298
DB 503 GCAGGCGCCCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 562
OY 299 agcagggcccaacagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 358
DB 563 ACCCGGAGAGTTCATTCACGCGGCTCGATGCCGCTTCTTGAACCCCTCCGAGGGCT 622
OY 359 ggtatgcctcgcgggttcctcatgtgcaagtcctggtatgagagagagagagag 418
DB 623 GGTGGCCGTTCTCTGATCAACATCTGATGAGGAGCGCGCGCTTCAAGCGCAGACA 682
OY 419 accctaccag 443
DB 683 TGTGGGCAAGACGAGGAGGATCCC 707

RESULT 15
US-09-324-542-40
Sequence 40, Application US/09324542
Patent No. 6328978
GENERAL INFORMATION:
APPLICANT: Watson, James D.
APPLICANT: Tan, Paul L.J.
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: Methods and Compounds for the Treatment
TITLE OF INVENTION: of Immunologically-Mediated Skin Disorders
FILE REFERENCE: 11000.1007c1
CURRENT APPLICATION NUMBER: US/09/324,542
CURRENT FILING DATE: 1999-06-02
EARLIER APPLICATION NUMBER: US 08/997,080
EARLIER FILING DATE: 1997-12-23
NUMBER OF SEQ ID NOS: 194
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 40
LENGTH: 1211
TYPE: DNA

ORGANISM: Mycobacterium vaccae
US-09-324-542-40

Query Match 2.58; Score 37; DB 4; Length 1211;
Best Local Similarity 44.66; Pred. No. 0.84;
Matches 145; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

```
QY 119 acacgagctcactcccggtccgcagagccatccccaggaggaagagatcc 178
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 383 acacgagctgtaagccccccgctacaagggcccgacgtgacctacaagtggaga 442
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 179 tcatgtgcacacaagcttcgggccaaggtgcagcctcaggcctccaacatgagtlaca 238
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 443 ccttctgaccacagagctcccggtctgctgcagggcaaccgcggtcaagccgacg 502
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 239 tgacctggagtgcagactggagaagtctgtcagcagctggcgactgaacctggcgacctggc 298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 503 gcagcgccctgtcggtctgtcgatggcggttcggcgcgctgaacctggcgacctggc 562
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 299 agcagcgggcccaacagctgtgtgtccatcggcagagaacctggggcgctcactggggca 358
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 563 acccgagagcagttcatctacgcgggtctcgatgtccggtctcctgaacccctccgagggt 622
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 359 ggtatcgctctccgggttcacatgtgcagctcctgtgtatgacgaggtgaagactaacct 418
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 623 ggtggcgcttctgtacacatctgatgggtgaagccgcggcttcaaggccgacgaca 682
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 419 accctaccgagcgagtgcaacc 443
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 683 tgtgggcaagaccgaggatccc 707
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: May 25, 2002, 13:23:33
Job time: 10224 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2002, 13:23:34 ; Search time 61 Seconds
(without alignments)
199.009 Million cell updates/sec

Title: US-09-667-380a-2

Perfect score: 2759
Sequence: 1 MSCVLGGVPLGLFLVCGS.....SESLGTPROGKAFRIRAVRQ 497

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5
231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCtus_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfilest.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	355	12.9	266	2	US-08-773-368-4
2	355	12.9	266	3	US-09-199-887-4
3	352	12.8	219	2	US-08-773-368-3
4	352	12.8	219	3	US-09-199-887-3
5	322	11.7	270	2	US-08-773-368-1
6	322	11.7	270	3	US-09-199-887-1
7	273	9.9	220	5	PCT-US96-07709-30
8	273	9.9	248	5	PCT-US96-07709-25
9	269	9.7	204	5	PCT-US96-07709-33
10	256	9.3	204	1	US-08-419-414-9
11	256	9.3	204	1	US-08-614-935-2
12	256	9.3	204	1	US-09-130-287-2
13	244.5	8.9	440	1	US-07-930-686-12
14	244.5	8.9	440	2	US-08-460-998-12
15	240	8.7	204	1	US-08-614-935-1
16	240	8.7	204	3	US-09-130-287-1
17	237	8.6	203	1	US-08-419-414-8
18	237	8.6	203	1	US-08-614-935-3
19	237	8.6	203	3	US-09-130-287-3
20	236.5	8.6	424	1	US-08-419-414-2
21	227	8.2	151	1	US-08-614-935-28
22	227	8.2	151	3	US-09-130-287-28
23	226	8.2	205	5	US-08-450-944-5
24	226	8.2	205	5	PCT-US96-07709-5
25	226	8.2	221	5	US-08-450-944-2
26	226	8.2	221	5	PCT-US96-07709-2
27	225.5	8.2	137	2	US-07-857-224B-110

28	222.5	8.1	205	1	US-08-419-414-10	Sequence 10, Appl
29	222.5	8.1	205	1	US-08-614-935-6	Sequence 6, Appl
30	222.5	8.1	205	1	US-08-614-935-7	Sequence 7, Appl
31	222.5	8.1	205	3	US-09-130-287-6	Sequence 6, Appl
32	222.5	8.1	205	3	US-09-130-287-7	Sequence 7, Appl
33	222	8.0	454	1	US-07-930-686-10	Sequence 10, Appl
34	222	8.0	454	2	US-08-460-998-10	Sequence 10, Appl
35	217.5	7.9	205	3	US-08-614-935-4	Sequence 4, Appl
36	217.5	7.9	204	1	US-08-614-935-5	Sequence 5, Appl
37	217	7.9	204	3	US-09-130-287-5	Sequence 5, Appl
38	217	7.9	135	2	US-07-857-224B-103	Sequence 103, App
39	216	7.8	166	1	US-08-614-935-30	Sequence 30, Appl
40	216	7.8	166	3	US-09-130-287-30	Sequence 29, Appl
41	212.5	7.7	150	3	US-08-614-935-29	Sequence 29, Appl
42	212.5	7.7	150	3	US-09-130-287-29	Sequence 112, App
43	212.5	7.6	138	2	US-07-857-224B-111	Sequence 111, App
44	211	7.6	137	2	US-07-857-224B-111	
45	210.5	7.6				

ALIGNMENTS

RESULT 1
US-08-773-368-4
; Sequence 4, Application US/08773368
; Patent No. 5856130
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Golli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN PATHOGENESIS-RELATED
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/773,368
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy RJ
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0186 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 266 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1030053
; US-08-773-368-4
Query Match 12.9%; Score 355; DB 2; Length 266;

```

1      RESULT      2
2      US-09-199-887-4
3      : Sequence 4, Application US/09199887
4      : Patent No. 6071874
5      : GENERAL INFORMATION:
6      : APPLICANT: Bandman, Olga
7      : APPLICANT: Goll, Surya K.
8      : TITLE OF INVENTION: NOVEL HUMAN PATHOGENESIS-RELATED
9      : TITLE OF INVENTION: PROTEIN
10     : NUMBER OF SEQUENCES: 6
11     : CORRESPONDENCE ADDRESS:
12     : ADDRESSEE: Incyte Pharmaceuticals, Inc.
13     : STREET: 3174 Porter Drive
14     : CITY: Palo Alto
15     : STATE: CA
16     : COUNTRY: USA
17     : ZIP: 94304
18     : COMPUTER READABLE FORM:
19     : MEDIUM TYPE: Diskette
20     : COMPUTER: IBM Compatible
21     : OPERATING SYSTEM: DOS
22     : SOFTWARE: FastSeq for Windows Version 2.0
23     : CURRENT APPLICATION DATA:
24     : APPLICATION NUMBER: US/09/199,887
25     : FILING DATE:
26     : CLASSIFICATION:
27     : PRIOR APPLICATION DATA:
28     : APPLICATION NUMBER: 08/773,368
29     : FILING DATE:
30     : ATTORNEY/AGENT INFORMATION:
31     : NAME: Billings, Lucy RJ
32     : REGISTRATION NUMBER: 36,749
33     : REFERENCE/DOCKET NUMBER: PF-0186 US
34     : TELECOMMUNICATION INFORMATION:
35     : TELEPHONE: 415-855-0555
36     : TELEFAX: 415-845-4166
37     : TELEX:
38     : INFORMATION FOR SEQ ID NO: 4:
39     : SEQUENCE CHARACTERISTICS:
40     : LENGTH: 266 amino acids
41     : TYPE: amino acid
42     : STRANDEDNESS: single
43     : TOPOLOGY: linear
44     : IMMEDIATE SOURCE:
45     : LIBRARY: Genbank
46     : CLONE: 1030053
47     : US-09-199-887-4

```

```

Yy      25 LPNVLLBELLSKYOHNSHESVRAIRPRED-KEELIMLHNKLRGOVOPOASNMETWTMD 83
        | : : : : : | : : : : : | : : : : : | : : : : : |
Db      5 LATIWMWSEFSNYSHT--ANILPDIEDNEDEIKOCVRILHNFRSEVKPTASDMITMTWD 61
        | : : : : : | : : : : : | : : : : : | : : : : : |
Qy     84 DELESAAMASOCCITWEHG-----PTSU---LVISIGONLGAMGRSRSGFHVOS----W 131
        | : | | | : | : | | : | : | | : | : | | : | : | |
Db     62 PALAQIAAKMASNOCFSNINTRLKPPHKHLPNNTSLGENT---W-TGSVIFESVSALTWW 117
        | : : : : : | : : : : : | : : : : : | : : : : : |
Qy    132 YDEVKDYYPPYPSECNPWCPCERSCGPCTHYTQIYWATTNKIGCAVNTRKRTWGGEYWE 191
        |||::||| : : : : : | : : : : : | : : : : : | : : : : : |
Db    118 YDELIDYDF-----KTRICKRKCVGHHYQYVWADSYKGCACAVQFCPKYSGF-DALS 166
        | : : : : : | : : : : : | : : : : : | : : : : : |
Qy     192 NAVPEVCVSPKGN---WIGEAPYKNRGRCSCPSYSSGCCNNNICYREE 238
        | : : : : : | : : : : : | : : : : : | : : : : : |
Db     167 NGAHFTCNVTGEGNTPTW---PYKRGAFTCSACPNN--DKCIDNLICVNRQ 210
        | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT          3
US-08-773-368-3
Sequence 3, Application US/08773368
Patent No. 5856130
GENERAL INFORMATION:
Applicant: Bandman, Olga
Applicant: Goll, Surya K.
TITLE OF INVENTION: NOVEL HUMAN PATHOGENESIS-RELATED
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/773.368
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy RJ
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PE-0186 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 219 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 847722
US-08-773-368-3

Query Match      12.8%, Score 352; DB 2; Length 219;
Best Local Similarity 37.0%; Pred. No. 1,3e-24;
Matches 81; Conservative 31; Mismatches 67; Indels 40; Gaps 12.
```

QY 94 ASOCIMHG-----PTSL---LVSIGONIGAHMGCRSPGFHVS-----WYDEVKDYTP 141
DB 62 ASNCFSHNTRKLPKHLHPNFTSLGENI---W-TGSVPISVSALITNWDIIDIYNF- 116
QY 142 YPSECNWPCRCGPMCTHTYQIYVATNTKIGCAVNTCRKMTWGEWENAVFYCVNS 201
DB 117 -----KTRICKKVCCHTYQVWADSYKVCACVFCPKVSGF-DALSNGAHFICNNG 166
QY 202 PKGN---WIGEAPYKNGRCSECPSPSYGSGCRNNLCYRE 237
DB 167 PGNTPTW---PYKRGATCSACPNN--DKCLDNLGVND 199

RESULT 4

US-09-199-887-3
Sequence 3, Application US/09199887
Patent No. 6071874
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
TITLE OF INVENTION: NOVEL HUMAN PATHOGENESIS-RELATED
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/199,887
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/773,368
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy RJ
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0186 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 219 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 847722
US-09-199-887-3

Query Match 12.8%; Score 352; DB 3; Length 219;
Best Local Similarity 37.0%; Pred. No. 1.3e-24;
Matches 81; Conservative 31; Mismatches 67; Indels 40; Gaps 12;
QY 35 LSKYONHESRVRRAIPRD-KETIMLHNKIRGOVQPOASNMEYMTWDELEKSAANV 93
DB 5 VSNSTH---ANTLPDENEDFKDCVRIHNKRSEYKPTASDMLMTWDPALAOAKAN 61
QY 94 ASOCIMHG-----PTSL---LVSIGONIGAHMGCRSPGFHVS-----WYDEVKDYTP 141

DB 62 ASNCFSHNTRKLPKHLHPNFTSLGENI---W-TGSVPISVSALITNWDIIDIYNF- 116
QY 142 YPSECNWPCRCGPMCTHTYQIYVATNTKIGCAVNTCRKMTWGEWENAVFYCVNS 201
DB 117 -----KTRICKKVCCHTYQVWADSYKVCACVFCPKVSGF-DALSNGAHFICNNG 166
QY 202 PKGN---WIGEAPYKNGRCSECPSPSYGSGCRNNLCYRE 237
DB 167 PGNTPTW---PYKRGATCSACPNN--DKCLDNLGVND 199

RESULT 5

US-08-773-368-1
Sequence 1, Application US/08773368
Patent No. 5856130
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
TITLE OF INVENTION: NOVEL HUMAN PATHOGENESIS-RELATED
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/773,368
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy RJ
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0186 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 270 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: 1599164
US-08-773-368-1

Query Match 11.7%; Score 322; DB 2; Length 270;
Best Local Similarity 35.6%; Pred. No. 1e-21;
Matches 84; Conservative 27; Mismatches 105; Indels 20; Gaps 8;
QY 50 AIPREDKEEILMLHNKIRGOVQPOASNMEYMTWDELEKSAANASOCIMHGPTSLVS 109
DB 27 AITDEKRLMVLNHLNLYRAQVSPFASDMLHMRWDELAFAKAYARQ--XRXGHNKERGR 84
QY 110 IQONIGAHMGKRRSOFHVSQWYDEVKDYTPYPSNCPWPCRCGPMCTHTYQIYVAT 169
DB 85 RGENLFAITDEGMDYPLAEWHEHREHYNS-AAATCSB-----GOMCHITLYVNAK 136
QY 170 TTKIGCAVNTCRKMTWGEWENAVFYCVNSPKGNWIGEADYKNGRCSECPSPSYGSGC 229

Db 137 TERIGCGSHFEKLO--GVEETINIELVANCPEPPGNVKGKRPQEGTCCSPSGY--HC 192
QY 230 RNNLCYRETYTPKPEDEMEVEETAPPEENHVLQPRVW--RPRPKTSAYN 282
Db 193 KNSLC--EPISGPDADLPYLVEAFSPRATED--SDSRKMGAGPDKPSVSGLN 245

RESULT 6

US-09-199-887-1
Sequence 1, Application US/09199887
Patent No. 6071874
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Goll, Surya K.
TITLE OF INVENTION: NOVEL HUMAN PATHOGENESIS-RELATED
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/199,887
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/773,368
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy RJ
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PP-0186 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 270 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: 1599164
US-09-199-887-1

Query Match 11.7%; Score 322; DB 3; Length 270;
Best Local Similarity 35.6%; Pred. No. 1e-21;
Matches 84; Conservative 27; Mismatches 105; Indels 20; Gaps 8;

QY 50 AIRDEKEEILMLHNLKLGQVPOASNMEXMTDDDELEKSAAMASQCIWEHGPTSLVLS 109
Db 27 ALTDEKRLMVELHNLRYQVSTASDMLHMRDELAFAPAKAYARQ--XRXGHNKERGR 84
QY 110 ICONLGAHNGRGRSPGFHVQSWYDEVKDYTPYPSHCNFWCPEKSGPACTHYTOIYVAT 169
Db 85 KENLEFAITDEGDVPLAMEEWHHEREHYNLS--AATCSP-----GQMGHITYVYMAK 136
QY 170 TFKIGCAVNTCKRMVWGVWVMAVYFVGNYSKGNWIGEAIFYKNGRPSGSECPSPSYGSGC 229
Db 137 TFRIGGSHFEKLO--GVEETINIELVANCPEPPGNVKGKRPQEGTCCSPSGY--HC 192
QY 230 RNNLCYRETYTPKPEDEMEVEETAPPEENHVLQPRVW--RPRPKTSAYN 282

Db 193 KNSLC--EPISGPDADLPYLVEAFSPRATED--SDSRKMGAGPDKPSVSGLN 245

RESULT 7

PCT-US96-07709-30
Sequence 30, Application PC/TUS9607709
GENERAL INFORMATION:
APPLICANT: Tripp, Cynthia A.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: PARASITIC HELMINTH VENOM ALLERGEN
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/07709
FILING DATE: 23-MAY-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-30-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-0223
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 220 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US96-07709-30

Query Match 9.9%; Score 273; DB 5; Length 220;
Best Local Similarity 29.5%; Pred. No. 2.5e-17;
Matches 70; Conservative 36; Mismatches 85; Indels 46; Gaps 10;

QY 16 LVCGSGGYLLP--NYTLLEELLSKYQHNSHSRVRRAIPREDKEILMLHNLKLR---GQV 70
Db 10 IYVAVAVGYNCPCGKLTALER-----KRIYQNNKRYSDILNKLNRNGTY 55
QY 71 OPOASNMEXMTDDDELEKSAAMASQCIWEHGPTSLVLSIGONLCAHW-----GRYRSP 124
Db 56 MPRGNMELRLNDCKLESSAORMANOCIFGHSPPROQREGVGENYAYVSSVSGLKTA 115
QY 125 GFHV--QSWYDEVKDYTPYPSHCNFWCPEKSGPACTHYTOIYVATNKKICCAVNT--C-- 180
Db 116 GYDACKSMWSELPRKYENPNNNMTW---KYAGQCVLHFYQWAMGKTKYKICGVAOTODG 172
QY 181 -RKMTYWEVWENAVFYCANSPPKGNWIGEAIFYKNGRPSGSECPSPSYGSGC--RNNLC 234
Db 173 GRTLV-----ICHYSPGNNMGEVITYGRNDCKYDKCYTKKLSRSGLC 218

RESULT 8

PCT-US96-07709-25
Sequence 25, Application PC/TUS9607709
GENERAL INFORMATION:
APPLICANT: Tripp, Cynthia A.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: PARASITIC HELMINTH VENOM ALLERGEN
TITLE OF INVENTION: ANTIGEN 5-LIKE GENES AND PROTEINS

NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/07709
FILING DATE: 23-MAY-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-30-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 248 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US96-07709-25

Query Match 9.9%, Score 273; DB 5; Length 248;
Best Local Similarity 29.5%, Pred. No. 2,9e-17;
Matches 70; Conservative 36; Mismatches 85; Indels 46; Gaps 10;

QY 16 LVGSGSYLLP--NTLLELLSKYOHNSHRSVRRAIPREDKEELIMLHNKLR---GOV 70
DB 38 IVAVVTGTCNCPGKITALER-----KKYGGNNKRYSDLLNGKLNRRNTY 83
QY 71 QPQASNNMEYTWDELEKSAAMAASOCIMEHPTSLVSIIGNLGAHW-----GRYSP 124
DB 84 MPRGNMELFLWDCLESSAQRMANOCIFGHSRPROREGEVYAYWSVSVEELKRTA 143
QY 125 GFHV-QSVYDEVKDYTPYPSSECPNCPERCSPMCTHTYQIYMATNKGAVNT-C-- 180
DB 144 GTDAGKSWSELPLKYENPNPNNMTW---KVAGGVLFHTOMAMGKTYKIGCVATQCDG 200
QY 181 -RKMTWGEVWENAVFYCANTSPKGNMIGEPYKNGRCPCECPSPSYGGSC--RNNIC 234
DB 201 GRTLLV-----ICHYSPGNNWGEVITYRGPNCKVYDKCKYKCKLSKSLG 246

RESULT 9
PCT-US96-07709-33
Sequence 33, Application PC/TUS9607709
GENERAL INFORMATION:
APPLICANT: tripp, Cynthia A.
APPLICANT: Wisniewski, Nancy
TITLE OF INVENTION: PARASITIC HELMINTH VENOM ALLERGEN
TITLE OF INVENTION: ANTIGEN 5-LIKE GENES AND PROTEINS
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/07709
FILING DATE: 23-MAY-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-30-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 204 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US96-07709-33

Query Match 9.7%, Score 269; DB 5; Length 204;
Best Local Similarity 31.0%, Pred. No. 5.2e-17;
Matches 63; Conservative 34; Mismatches 76; Indels 30; Gaps 8;

QY 48 RRAIPREDKEELIMLHNKLR---GOVOPASNNMEYTWDELEKSAAMAASOCIMEHPT 104
DB 14 KKIYGGNNKRYSDLLNGKLNRRNTYMPRGKMLELWDCLESSAQRMANOCIFGHSR 73
QY 105 SLVSIIGNLGAHW-----GRYSPGFHV-QSVYDEVKDYTPYPSSECPNCPERCSP 157
DB 74 QORGEVGENYAYWSSVSYGLKKTAGTDAGKSWSELPLKYENPNPNNMTW---KVAGQ 130
QY 158 MCHYQIYMATNKGAVNT-C---RKMTWGEVWENAVFYCANTSPKGNMIGEPYKNGR 213
DB 131 GVLHTOMAMGKTYKIGCVATQCDGGRLLV-----ICHYSPGNNWGEVITY 179
QY 214 NGRPCSECPSPSYGGSC--RNNIC 234
DB 180 RGNPKVDKDYTKKCKLSKSLG 202

RESULT 10
US-08-419-414-9
Sequence 9, Application US/08419414
Patent No. 5753787
GENERAL INFORMATION:
APPLICANT: Hawdon, John M.
APPLICANT: Hotez, Peter J.
APPLICANT: Jones, Brian F.
TITLE OF INVENTION: Hookworm Vaccine
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/419,414
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284

REFERENCE/DOCKET NUMBER: YU113
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8795
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 204 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Vespula vulgaris
US-08-419-414-9

Query Match 9.3%; Score 256; DB 1; Length 204;
Best Local Similarity 30.5%; Pred. No. 8.1e-16;
Matches 71; Conservative 33; Mismatches 73; Indels 56; Gaps 12;

QY 1 MSCVIGVIRLGLFLVCGSGYLIPNVLLELLSKYQHNSHSRVRRAIRREDEEIL 60
DB 6 IKCLKGV-----HTAC-KYSLKPNCG--NKVYVS-----GLTKQKODIL 45
QY 61 MLNKLKGV-----OPASNMEYTWDELEKSAAMASOCIEHGPSTSLV 108
DB 46 KEHNDPRKRIANGLETGRNPGPQPPAKNMKNLVNDELAYVAQVWANCQYGHDTCDVA 105
QY 109 --SIGONT--GAHNGRRSPGFHVSQWYDEVDKYTPYPSKCPMPCPERCSP--MCT 160
DB 106 KYQVQGNVALGSIATAKDDPKYLVKMEDEVDKY-----NP--KKKFSGNDLKTG 155
QY 161 HTQIVMATTKIGCAVNTCRKMTWGEVWENAVFYFCNTSPKGNMGEAPYK 213
DB 156 HTQIVMATTKIGCAVNTCRKMTWGEVWENAVFYFCNTSPKGNMGEAPYK 213
156 HTQIVMATTKIGCAVNTCRKMTWGEVWENAVFYFCNTSPKGNMGEAPYK 213
156 HTQIVMATTKIGCAVNTCRKMTWGEVWENAVFYFCNTSPKGNMGEAPYK 213

RESULT 11
US-08-614-935-2

Sequence 2, Application US/08614935
Patent No. 5804201
GENERAL INFORMATION:

APPLICANT: King, Te P.
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES OF VESPID
NUMBER OF SEQUENCES: 81
CORRESPONDENCE ADDRESS:
ADDRESSSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/614,935
FILING DATE: 11-MAR-1996
CLASSIFICATION: 436

ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-156
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 204 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM: Vespula vulgaris
US-08-614-935-2

Query Match 9.3%; Score 256; DB 1; Length 204;
Best Local Similarity 30.5%; Pred. No. 8.1e-16;
Matches 71; Conservative 33; Mismatches 73; Indels 56; Gaps 12;

QY 1 MSCVIGVIRLGLFLVCGSGYLIPNVLLELLSKYQHNSHSRVRRAIRREDEEIL 60
DB 6 IKCLKGV-----HTAC-KYSLKPNCG--NKVYVS-----GLTKQKODIL 45
QY 61 MLNKLKGV-----OPASNMEYTWDELEKSAAMASOCIEHGPSTSLV 108
DB 46 KEHNDPRKRIANGLETGRNPGPQPPAKNMKNLVNDELAYVAQVWANCQYGHDTCDVA 105
QY 109 --SIGONT--GAHNGRRSPGFHVSQWYDEVDKYTPYPSKCPMPCPERCSP--MCT 160
DB 106 KYQVQGNVALGSIATAKDDPKYLVKMEDEVDKY-----NP--KKKFSGNDLKTG 155
QY 161 HTQIVMATTKIGCAVNTCRKMTWGEVWENAVFYFCNTSPKGNMGEAPYK 213
DB 156 HTQIVMATTKIGCAVNTCRKMTWGEVWENAVFYFCNTSPKGNMGEAPYK 213
156 HTQIVMATTKIGCAVNTCRKMTWGEVWENAVFYFCNTSPKGNMGEAPYK 213
156 HTQIVMATTKIGCAVNTCRKMTWGEVWENAVFYFCNTSPKGNMGEAPYK 213

RESULT 12
US-09-130-287-2

Sequence 2, Application US/09130287
Patent No. 6106844
GENERAL INFORMATION:

APPLICANT: King, Te P.
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES OF VESPID
NUMBER OF SEQUENCES: 81
CORRESPONDENCE ADDRESS:
ADDRESSSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/130,287
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/614,935
FILING DATE: 11-MAR-1996
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-156 DIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 204 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEetical: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM: Vespula vulgaris
US-09-130-287-2

Query Match 9.3%; Score 256; DB 3; Length 204;
Best Local Similarity 30.5%; Pred. No. 8.1e-16;
Matches 71; Conservative 33; Mismatches 73; Indels 56; Gaps 12;

QY 1 MSCVLGVIPILGLFLVCGSGYLLPNTLLLELLSKYOHNSHVRRAIPREDKEILL 60
DB 6 IKCLKGV-----HTAC-KYGLKPNCG--NKVVS-----GLTKOEKODILL 45
QY 61 MHNLRGV-----QPOASNMEYMTWDELEKSAANASOCIWEHPTSLV 108
DB 46 KEHNDEROKIARLETRGNPGPPAKNKNLVWMDLAVYAVQVWANCQYGHDTCRDVA 105
QY 109 --SIGONT--GAHMGKRYSPGFHVQSWYDEVKDYTPYRSECPNWCPCRGSGP--MCT 160
DB 106 KVOVGONVALITSTAKYIDPYKLVKMWDEYKDY-----NP--KKKFSNDLKLNG 155
QY 161 HTQIVATNTKICAVNTCRKMTWGEVWENAVFYCNYSKGNMIGEARPYK 213
DB 156 HTQVWVWANTKEVGGC-----SIKYIOEKMKH-KYLVCHVPSGNMNEELYQ 202

RESULT 13

US-07-930-686-12
Sequence 12, Application US/07930686
Patent No. 5525508
GENERAL INFORMATION:
APPLICANT: Sharp, Phillip J
APPLICANT: Magland, Barry M
APPLICANT: Cobon, Gary S
TITLE OF INVENTION: Nematode Vaccine
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley and Lardner
STREET: suite 500, 1800 Diagonal Road
CITY: Alexandria
STATE: Virginia
COUNTRY: United States of America
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/930,686
FILING DATE: 19921006
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK4486
FILING DATE: 06-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU92/00040
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A
REGISTRATION NUMBER: 29,768
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109

TELEX: 899149
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 440 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-930-686-12

Query Match 8.9%; Score 244.5; DB 1; Length 440;
Best Local Similarity 29.7%; Pred. No. 2.8e-14;
Matches 73; Conservative 32; Mismatches 86; Indels 55; Gaps 10;

QY 39 OHNESHRRRAIPREDKEILLMHNKRGVQPOASNMEYMTWDELEKSAANASOCI 98
DB 32 KNEHYRSIIAKG-----QAKNKL-GGFAPKARMLKVYDDEVEANFAVAKCK 80
QY 99 WEHPTSLVLSIGONT--GAHMGKRYSPGFHVQSWYDEVKDYTPYRSECPNWCPCRG 154
DB 81 FEHDPEORRYWGNLMMLGTYSTESAKLSVQAWYMWELKMGVDPDENILTMVEYFDG 140
QY 155 SGPMCHHTQIVATNTKICAVNTCRKMTWGEVWENAVFYCNYSKGNMIGEARPYKN 214
DB 141 VG-----HTQVWVWANTKEVGGC-----VACETNPAGNRINHYYDI 186
QY 215 GRPCS--ECPPSYGSGCRNNLCYRETY-----TPKPEDEMEVETAP----- 256
DB 187 GDCPTDEDC-----QCCTCCKSDKALCIPGYTTVMPTPE--KPTTPKLYHPGGM 238
QY 257 TPEENH 262
DB 239 CPENNN 244

RESULT 14

US-08-460-998-12
Sequence 12, Application US/08460998
Patent No. 5942413
GENERAL INFORMATION:
APPLICANT: Sharp, Phillip J
APPLICANT: Magland, Barry M
APPLICANT: Cobon, Gary S
TITLE OF INVENTION: Nematode Vaccine
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley and Lardner
STREET: suite 500, 3000 K Street, NW
CITY: Washington
STATE: DC
COUNTRY: United States of America
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,998
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/930,686
FILING DATE: 06-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK4486
FILING DATE: 06-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU92/00040
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A
REGISTRATION NUMBER: 29,768
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 440 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-460-998-12

Query Match 8.9%; Score 244.5; DB 2; Length 440;
Best Local Similarity 29.7%; Pred. No. 2.8e-14;
Matches 73; Conservative 32; Mismatches 86; Indels 55; Gaps 10;

QY 39 OHNESHRYRAIPREKEELIMHNKLRGOVQASNMETWDELESAAMASQCI 98
DB 32 KHNEYRSLIAKG-----QAKNKL-GGFAFKARMLKVCYDCEVANTAAVAKCK 80
QY 99 WEHPTSLVLSIGNTL-----GAHWGRYRSPGFHVOSWYDEKDYTPYPSQCNPCPERC 154
DB 81 FEHDPPEQRWYQONLMGLSTVSKTESAKLSVQAMYTELKFVDENLITVEYFDG 140
QY 155 SCMCCTHYTQIYVATNKGICAVNTCRKMTWGEWENAVYFVCSNPKGNWIGEPYKN 214
DB 141 VG---HYTQVAMQSSDKICAVEMCPTMTL-----VACEYVAPNRIRHYYIDI 186
QY 215 GRCS---ECPPSYGSGCRNNLCYRETY-----TPKETDEMEVEVAP----- 256
DB 187 GDCCTTDEDC-----QCTGCTCSKDEALCIPGYTTPMPTE--KPTTPKLYHNGM 238
QY 257 IPEENH 262
DB 239 CPENNN 244

RESULT 15
US-08-614-935-1
Sequence 1, Application US/08614935
Patent No. 5804201
GENERAL INFORMATION:
APPLICANT: King, Te P.
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES OF VESPID
NUMBER OF SEQUENCES: 81
CORRESPONDENCE ADDRESS:
ADDRESSSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
FLOOR
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/614,935
FILING DATE: 11-MAR-1996
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-156
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 204 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORGANISM: Vespula maculifrons
US-08-614-935-1

Query Match 8.7%; Score 240; DB 1; Length 204;
Best Local Similarity 29.6%; Pred. No. 2.4e-14;
Matches 68; Conservative 27; Mismatches 85; Indels 50; Gaps 9;

QY 1 MSCVLAGVTPGLLEFLVCGSQGYLLPNVTLLELLSKYOHNESHRYRAIPREDKEIL 60
DB 6 IKCLKGV-----HTAC-KYGSILKPN-----GNKKVSYGLTKQEKODIL 45
QY 61 MLHKKLGGV-----QPASNMEYTWDELESAAMASQCIWEHPTSLV 108
DB 46 KEHNDPQKIAKGLKLETRGNPQPQPPAKNMKLVMSDELATYQVWANOQYGHDTCDVA 105
QY 109 --SIGNTL-----GAHWGRYRSPGFHVOSWYDEKDYTPYPSQCNPCPERCSPMCTHYT 163
DB 106 KYQGVQVVALTGSTAANYNDVKLVKMEDEVKDYNPKKKFSENNFL-----KIGHT 158
QY 164 QIYVATNKGICAVNTCRKMTWGEWENAVYFVCSNPKGNWIGEPYK 213
DB 159 QMWVANTRKEVGG-----SIKIQENMKH-YLVCSNPGSGNPQNEELYQ 202

Search completed: May 25, 2002, 14:28:50
Job time: 3916 sec

Tue May 28 16:10:04 2002

us-09-667-380a-2.rpt

Page 2

C:Accession: J04131
R:Murphy, E.V.; Zhang, Y.; Zhu, W.; Biggs, J.
Gene 159, 131-135, 1995
A:Title: The human glioma pathogenesis-related protein is structurally related to plant
A:Reference number: J04131; MUID:9531646
A:Molecule type: mRNA
A:Residues: 1-219 <MUD>
A:Cross-references: GB:016307; NID:91100927; PID:AAA2731.1; PID:g847722
A:Experimental source: brain tumor
A:Gene: GDB:GILPR
A:Superfamily: yellowjacket venom allergen antigen 5
C:Keywords: brain

Query Match
Best Local Similarity 12.8%; Score 352; DB 2; Length 219;
Matches 81; Conservative 31; Mismatches 67; Indels 40; Gaps 12;

QY 35 LSKYQHSHSRVRAIPRED-KEEILMLHNKLGVOQOASNMETWDELKSAAM 93
DB 5 VSNYSR---ANILPDIEDNEDPINDCVRIHNRKSEVKTSDMLYTPDPAIAQAKAM 61
QY 94 ASQCIWEG---PTSL---LVISQNLGAMHGRYRSPGFHYOS---WYDEVKDYTP 141
DB 62 ASNCOFSHNTRLKPPHKLHFNFTSLGENT---W-TGSVPFVSASAITWYDEIDYNF- 116
QY 142 YPSECNWCPCRGSGMCTHYQIYMATNKIGCAVNTCKMTWCEWENAVYFCNYS 201
DB 117 -----KTRICKKVGCGHYTYVYVADSYKVCANVQFCPKVSGP-DALSGAHFICNG 166
QY 202 PKNN---WIGFAPYKNGPCSECPSPYSGSCNNLCYRE 237
DB 167 PGGNPTW---PYKRGATSCACPNV-DKCLDNLCVND 199

RESULT 3
A44609
acidic epididymal glycoprotein precursor - rat
N:Alternate names: sperm-coating glycoprotein
C:Species: Rattus norvegicus (Norway rat)
C>Date: 30-Jun-1988 #sequence (Morway rat)
R:Accession: A40918; MUID:9531646
R:Character: N.J.; Joseph, D.R.; Wilson, E.M.; French, F.S.
A:Title: Molecular cloning of the cDNA for androgen-dependent sperm-coating glycoprotein
A:Reference number: A44609; MUID:87053995
A:Accession: A40918; MUID:89039913
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-246 <CHN>
A:Cross-references: GB:M31173; NID:9202772; PID:AA859716.1; PID:9202773
R:Brooks, D.E.; Means, A.R.; Wright, E.J.; Singh, S.P.; Taver, K.K.
A:Title: Molecular cloning of the cDNA for androgen-dependent sperm-coating glycoprotein
A:Reference number: A44609; MUID:87053995
A:Accession: A40918; MUID:89039913
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-246 <CHN>
A:Cross-references: GB:X04643; NID:956112; PID:CAA28304.1; PID:956113
C:Superfamily: cysteine-rich secretory protein 1
C:Keywords: glycoprotein; sperm
F:1-19/Domain: signal sequence
F:20-246/Product: acidic epididymal glycoprotein
Query Match
Best Local Similarity 10.9%; Score 301.5; DB 2; Length 246;
Matches 81; Conservative 51; Mismatches 90; Indels 81; Gaps 15;

DB 5 LVLEFLAA-----VLP-PSLLDPTDDEMRDLENSTKLVS-----QEEIKHNOQLRT 54
QY 70 VPOASNMETWDELKSAAMASOCIWEHGP---TSLVSGNTL-----GAHGRY 121
DB 55 VSPSGDILLVEMDHDVAVNAOKMANRCTYNHSPLOHRTTTLT----- 114
QY 122 RSPGFHYOSMYDEVKDYTPYPSNCPNWCPCRGSGMCTHYQIYMATNKIGCAVNTCR 181
DB 115 -----IDWIDESLIDFVFGG-----PKRV-GVAKGHITVYVNSFLVACVACCP 160
QY 182 KMTWGEWENAVYFCNYSFGKNNIGE--APYKNGRPGCECPSPYSGSCNNLCYREET 239
DB 161 DQPL-----KFFYVCHCPGQNVGRVLSPTLEGFCDSCP-----GNCEDGLCTNSCE 209
QY 240 YTPKPEIDENNEVEYAPIDERNHVMLOPRVMPRTKPKTSAVNVYQVYRCDTR-MKDRG 298
DB 210 Y-----EDNY-----SNCGDLKMYNSCDPLLEKGC 235
QY 299 KGS 301
DB 236 RAS 238

RESULT 4
A49202
cysteine-rich secretory protein-1 - mouse
N:Alternate names: CRISP-1
C:Species: Mus musculus (house mouse)
C>Date: 19-Dec-1993 #sequence (house mouse)
R:Accession: A49202
R:Haendler, B.; Kratzschmar, J.; Theuring, F.; Schlemming, W.D.
A:Title: Transcripts for cysteine-rich secretory protein-1 (CRISP-1; DE/ABC) and the
A:Reference number: A49202; MUID:93307144
A:Accession: A49202; MUID:93307144
A:Status: preliminary
A:Molecule type: nucleic acid
A:Cross-references: GB:I05559; NID:9309190; PID:AAA37460.1; PID:9309191
A:Experimental source: NMRI, epididymis, salivary gland
A:Note: Sequence extracted from NCBI backbone (NCBI:134675, NCBI:134676)
C:Superfamily: cysteine-rich secretory protein 1
Query Match
Best Local Similarity 10.8%; Score 298; DB 2; Length 244;
Matches 77; Conservative 35; Mismatches 87; Indels 42; Gaps 11;

QY 11 LGLELVGSGGILPNVTLLEELLSKYOHNSHVRRAIPREDKEEILMLHNKLGVO 70
DB 5 LVLEFLAA-----VLP-PSLLD-SSOENLEKLTITKMSV-----QEEIKHNOQLRT 54
QY 71 OPOASNMETWDELKSAAMASOCIWEHGP---TSLVSGNTL-----GAHGRY 121
DB 54 SPSSGDLKMEWMDVAVNAOQADKCFESHPSITLTYMLRGENTLMSLSWS-- 111
QY 123 SPGRHYOSMYDEVKDYTPYPSNCPNWCPCRGSGMCTHYQIYMATNKIGCAVNTCR 182
DB 112 -----AIOGVNEXKDLT-----DVGPKQDPSYVG-----HYTVVNSIFVOAGVACCP 159
QY 183 KMTWGEWENAVYFCNYSFGKNNIGE--APYKNGRPGCECPSPYSGSCNNLCYREET 240
DB 160 NPL-----RYTVVCHCPGQNVGRVLSPTLEGFCDSCP-----GNCEDGLCTNSCE 209
QY 241 T 241
DB 213 T 213

RESULT 5
S68691
neutrophil granules matrix glycoprotein SGP28 precursor - human

C:Species: Homo sapiens (man)
 C:Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 17-Nov-2000
 C:Accession: S68691; S74313; S68683
 R:Kjeldsen, L.; Cowland, J.B.; Johnsen, A.H.; Borregaard, N.
 FEBS Lett. 380, 246-250, 1996
 A:Title: SGP28, a novel matrix glycoprotein in specific granules of human neutrophils
 A:Reference number: S68691; MUID:96186934
 A:Accession: S68691
 A:Molecule type: mRNA
 A:Residues: 1-243 <KJ>
 A:Cross-references: EMBL:X94323; NID:g1213612; PIDN:CAA63984.1; PID:g1213613
 A:Accession: S74313
 A:Molecule type: protein
 A:Residues: 33-83;96-143;165-217;221-226 <KJ>
 R:Kraetzschmar, J.; Haendler, B.; Eberspaecher, U.; Roosterman, D.; Donner, P.; Schleunz, J.; Blochman, J. 236, 827-836, 1996
 Eur. J. Biochem. 236, 827-836, 1996
 A:Title: The human cysteine-rich secretory protein (CRISP) family. Primary structure and function
 A:Reference number: S68681; MUID:96270732
 A:Accession: S68683
 A:Molecule type: mRNA
 A:Residues: 1-105;107-245 <KRA>
 A:Cross-references: EMBL:X95240; NID:g1262818; PIDN:CAA64527.1; PID:g1262819
 A:Genetics:
 A:Gene: SGP28
 C:Superfamily: cysteine-rich secretory protein 1
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-245/Product: neutrophil granules matrix glycoprotein SGP28 #status predicted <MAT>

Query Match 10.7%; Score 295.5; DB 2; Length 245;
 Best Local Similarity 31.0%; Pred. No. 1.9e-15;
 Matches 75; Conservative 29; Mismatches 91; Indels 47; Gaps 10;

```

QY 13 LLEVLGSGGYLLPVNTLLELLSKYONHSHSVRAI--PRDKEELMLHKLKQ 69
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 7 LLEVLGAG-----LLSPFANEDKDPFTALLTTOLOVOREIVKHNELRA 52

QY 70 VOPQASNMETWDELEKSAAMASQCIWEH-GPTSLIVSI--GONL-----GAHWGRY 121
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 53 VSPFANMLKMEWKNEMAKNAQKMAQCNVRIHNPDRMTSLKGENLYMSSAPSSWSQ 111

QY 122 RSPGFVQSYDEKDYTPYSECPNCPERCSPGMCPTHYQIYVATTNKIGCAVNTIC 181
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 112 ----AIOSEFEDYNDFF----GVGPKPNAVVG----HYIQVWVSYLVGCGNAYCP 158

QY 182 KMTWGEWENAVYFCVNTSPKGNWIG--APYKNGRGPSECPSPYGGSCRNLLCYREE 239
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 159 NOKVL-----KYYVYQYCPAGNMANRLVYVEQGAPCASCPCDNDGLCTGCKYEDL 212

QY 240 YT 241
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 213 YS 214
  
```

RESULT 6
 B33329
 Cysteine-rich secretory protein 2 type I precursor - human
 N:Alternate names: testis-specific protein
 C:Species: Homo sapiens (man)
 C:Date: 09-Mar-1990 #sequence_revision 09-Mar-1990 #text_change 20-Jun-2000
 C:Accession: B33329; S68682
 R:Kaashara, M.; Gutknecht, J.; Brew, K.; Spurr, N.; Goodfellow, P.N.
 Genomics 5, 527-534, 1989
 A:Title: Cloning and mapping of a testis-specific gene with sequence similarity to a sperm
 A:Reference number: A33329; MUID:90129048
 A:Accession: B33329
 A:Molecule type: mRNA
 A:Residues: 1-243 <KAS>
 A:Cross-references: GB:M25532; NID:g339882; PIDN:AAA61220.1; PID:g339883
 R:Kraetzschmar, J.; Haendler, B.; Eberspaecher, U.; Roosterman, D.; Donner, P.; Schleunz, J.; Blochman, J. 236, 827-836, 1996

A:Title: The human cysteine-rich secretory protein (CRISP) family. Primary structure
 A:Reference number: S68681; MUID:96270732
 A:Accession: S68682
 A:Molecule type: mRNA
 A:Residues: 1-243 <KRA>
 A:Cross-references: EMBL:X95239; NID:g1262816; PIDN:CAA64526.1; PID:g1262817
 C:Genetics:
 A:Gene: GDB:TPX1
 A:Cross-references: GDB:120760; OMIM:187430
 A:Map position: 6p21-qter
 C:Superfamily: cysteine-rich secretory protein 1
 F:1-20/Domain: signal sequence #status predicted <SIG>
 F:21-243/Product: cysteine-rich secretory protein 2 type I #status predicted <MAT>

Query Match 10.2%; Score 281; DB 2; Length 243;
 Best Local Similarity 26.3%; Pred. No. 2.5e-14;
 Matches 83; Conservative 33; Mismatches 83; Indels 116; Gaps 14;

```

QY 24 LLEVLGSGGYLLPVNTLLELLSKYONHSHSVRAI--PRDKEELMLHKLKQ 68
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 3 LLEVLGAG-----SLPFGKDPFTALLTTOLOVOREIVKHNELRA 49

QY 69 VOPQASNMETWDELEKSAAMASQCIWEH-GPTSLIVSI--GONLGAHWGRYRSP 124
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 50 AVSPFANMLKMEWKNEMAKNAQKMAQCNVRIHNPDRMTSLKGENLYMSSAPSSWSQ 103

QY 125 ----GFVQSYDEKDYTPYSECPNCPERCSPGMCPTHYQIYVATTNKIGCAVNTIC 180
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 104 PTSSSAIQSYDEIDLFVY----GVGPKSPNAVVG----HYIQVWVSYLVGCGIATC 155

QY 181 KMTWGEWENAVYFCVNTSPKGNWIG--EAPYKNGRGPSECPSPYGGSCRNLLCYREE 238
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 156 PNOD-----SLKYYVYQYCPAGNMANRLVYVEQGAPCASCPCDNDGLCTGCKYEDL 201

QY 239 TYTPKPEIDENNEVEFAPPEENHVLQPRVYRPTKERTSAVNYVTVQVRCDF 292
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 202 -----TMSQYODLLSNCDLKNFTAG 222

QY 293 ----KMKDRKGSTC 303
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 223 CEHELKCK-ATC 236
  
```

RESULT 7
 T16415
 Hypothetical protein F48B8.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T16415
 R:Kirsten, J.
 Submitted to the EMBL Data Library, March 1995
 A:Description: The sequence of C. elegans cosmid F48B8.
 A:Reference number: S59413
 A:Accession: T16415
 A:Molecule type: translated from GB/EMBL/DBJ
 A:Residues: 1-312 <KIR>
 A:Cross-references: EMBL:U23514; NID:g746484; PID:g746485; PIDN:AA64538.1; CESP:F48E
 C:Genetics:
 A:Gene: CESP:F48E8.1
 A:Introns: 28/1; 94/3; 153/3; 174/3; 226/3

Query Match 9.7%; Score 269; DB 2; Length 312;
 Best Local Similarity 28.4%; Pred. No. 2.8e-13;
 Matches 83; Conservative 36; Mismatches 93; Indels 80; Gaps 17;

```

QY 21 QGYLLPVNTLLELLSKYONHSHSVRAI--PRDKEELMLHKLKQ 70
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 55 RGYFPFHFQSDGLSRSEHPNDEYLRKWTTHENRRRVV----- 96
  
```

Query 71 QPQASNMEYMTWDELLEKSAAMASOCIEHGPISLVSIGONICAHMGRYRSPGFHVOS 130
 Db 97 --ASDMMMLYMSDELASQRIHDTGCFRHSRGR--INVENTIA--APYSNDIAISI 149
 QY 131 WYDEKDYTPYPSRCNWP-----ERCSGPMCTHYTOIYMATNKGAVNCRKMT- 184
 Db 150 WFNVEH-----NTRCCNHAYKHC-----CGHYVQVWMAKTNLVGCGFSRCRYOG 195
 QY 185 WGEWENAVYVCNYSRKNMI-----GE-----APYKRGRCSECPSPSYGSSCR 230
 Db 196 WVGCHGRNV--FVCHYNDQGTVEVYARQYAMPATWSAGDNGK--CSNCPAN--APACY 251
 QY 231 NNTCYREYTPKPEDEMEVEETAPIDENHNLQPRVAPRTKPKTSAYN 282
 Db 252 QGLCYMKPKNEAPTTTE--STTSTTTE-----PTTCPEPEPEAGADN 296

RESULT 8
 JE0204
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 21-Jul-2000
 C:Accession: JE0204
 R:Maeda, T.; Sakashita, M.; Ohba, Y.; Nakanishi, Y.
 Biochem. Biophys. Res. Commun. 248, 140-146, 1998
 A>Title: Molecular cloning of the rat tpx-1 responsible for the interaction between spect
 A:Reference number: JE0204; MUID:98340864
 A:Accession: JE0204
 A:Molecule type: mRNA
 A:Residues: 1-243 <MAE>
 A:Cross-references: DDBJ:AB009662; NID:93374579; PIDN:BA022029.1; PID:93374580
 C:Comment: This protein functions as a cell adhesion protein for the association between
 C:Genetics:
 A:Map position: 17
 C:Superfamily: cysteine-rich secretory protein 1

Query Match 9.7%; Score 268.5; DB 2; Length 243;
 Best Local Similarity 31.7%; Pred. No. 2,3e-13;
 Matches 60; Conservative 29; Mismatches 65; Indels 35; Gaps 8;

QY 56 KEELMLHNKLRGVQVOPQASNMEYMTWDELLEKSAAMASOCIEHGPISLVSIGQ 112
 Db 38 QREITAKHNEIRROVSPGSGNLMKEMNVQAANAOKWANNICILEHSTEDRKINICGE 97
 QY 113 NL-----GAHMGRRYRSPGFHVOSWYDEKDYTPYPSRCNWPCEPSCMCTHYTOIYW 167
 Db 98 NLVYSTDPTSMRTV-----IQSWYENENFVGVGAK-----PNSAVG-----HYTDLW 142
 QY 168 ATTAKICAVNTCKKMTWGEWENAVYVCNYSRKNMI--GEAPYKRGRCSECPSPSY 225
 Db 143 YSSKRVKCGVAYCPNODTL-----KTYVCHYCHCPMGNNVKKSTPYHQSTPCASCP--- 193
 QY 226 GSGCRNNLC 234
 Db 194 --NNCDNGLC 201

RESULT 9
 A33329
 C:Species: Mus musculus (house mouse)
 C>Date: 09-Mar-1990 #sequence_revision 09-Mar-1990 #text_change 29-Sep-1999
 C:Accession: A33329
 R:Kasahara, M.; Gutknecht, J.; Brew, K.; Spurr, N.; Goodfellow, P.N.
 A>Title: Cloning and mapping of a testis-specific gene with sequence similarity to a spe
 A:Reference number: A33329; MUID:90129048
 A:Accession: A33329
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-243 <KAS>

A:Cross-references: GB:M25533; NID:9202126; PIDN:AAAA0472.1; PID:9202127
 C:Superfamily: cysteine-rich secretory protein 1

Query Match 9.7%; Score 267.5; DB 2; Length 243;
 Best Local Similarity 33.2%; Pred. No. 2.8e-13;
 Matches 63; Conservative 28; Mismatches 62; Indels 37; Gaps 9;

QY 56 KEELMLHNKLRGVQVOPQASNMEYMTWDELLEKSAAMASOCIEHGPISLVSIGQ 112
 Db 38 QREITAKHNEIRROVSPGSGNLMKEMNVQAANAOKWANNICILEHSTEDRKINICGE 97
 QY 113 NL-----GAHMGRRYRSP-----GFHVOSWYDEKDYTPYPSRCNWPCEPSCMCTHYTOIYW 167
 Db 98 NL-----YMSDPTLMVSIQSWYENEDFVYGAK-----PNSAVG-----HYTDLW 142
 QY 168 ATTAKICAVNTCKKMTWGEWENAVYVCNYSRKNMI--GEAPYKRGRCSECPSPSY 224
 Db 143 YSSKRVKCGVAYCPNODTL-----KTYVCHYCHCPMGNNVKKSTPYHQSTPCASCP--- 193
 QY 225 YGSGCRNNLC 234
 Db 194 --NNCENGLC 201

RESULT 10
 A44583
 C:Species: Vespa flavopilosa
 C>Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 11-Jan-2000
 C:Accession: A44583; B44522
 R:Hoffman, D.R.
 J. Allergy Clin. Immunol. 92, 707-716, 1993
 A>Title: Allergens in Hymenoptera venom XXV: the amino acid sequences of antigen 5 mo
 A:Reference number: A44583; MUID:9404316
 A:Accession: A44583
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-204 <HOP>
 C:Superfamily: yellowjacket venom allergen antigen 5

Query Match 9.2%; Score 253; DB 2; Length 204;
 Best Local Similarity 30.5%; Pred. No. 3e-12;
 Matches 71; Conservative 32; Mismatches 74; Indels 56; Gaps 12;

QY 1 MSCVIGGVPLGLFLVCGSGGILLPNTLLLELSKYOHNSHSRVRAIPREDKEIL 60
 Db 6 IKLAKGV-----HTAC--KYGSLKPNCG--NKVVVS-----GLTKQKQDIL 45
 QY 61 MLHNKLRGV-----OPQASNMEYMTWDELLEKSAAMASOCIEHGPISLVSIGQ 108
 Db 46 KEHNDPFRQIANGLETGPNQPPAKKMKNLWMDLAAYVAQWANOCCYCHDTCRDIA 105
 QY 109 --SIGONL-----GAHMGRRYRSPGFHVOSWYDEKDYTPYPSRCNWPCEPSCMCTHYTOIYW 167
 Db 106 KYQVGNVALGISTAKYDDPKYIKVMEDEYKDY-----NP--KKKSGNNFLKTG 155
 QY 161 HYTQIYMATNKGAVNCTCKKMTWGEWENAVYVCNYSRKNMI--GEAPYKRGRCSECPSPSY 213
 Db 156 HYTQVWMAKTNKVEGCG-----SIFIOEKMKH--YLVANCYGPSGNFQNEELYQ 202

RESULT 11
 B37350
 C:Species: Solenopsis invicta (red imported fire ant)
 C>Date: 01-Oct-1992 #sequence_revision 01-Oct-1992 #text_change 11-Jan-2000
 R:Hoffman, D.R.
 J. Allergy Clin. Immunol. 91, 71-78, 1993
 A>Title: Allergens in hymenoptera venom XXIV: the amino acid sequences of imported fl

RESULT 13
B37329

178 OY

QY 178 NTCRKMTVMGEVWENAVYFVCNYSPOKN-WIGLAPYKNGRPOSECPSPYGGSCRNNLC-234
::: ::::| | | | | :| :|

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2002, 14:28:55 ; Search time 63.23 seconds
(without alignments)
304.343 Million cell updates/sec

Title: us-09-667-380A-2

Perfect score: 2759
Sequence: 1 MGVGLGVPLGLFLVCGS.....SESLGTPDGKAFRIFAVRQ 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	355	12.9	266	1 GLIP_HUMAN	P48060 homo sapien
2	301.5	10.9	246	1 AEG_RAT	P12020 rattus norv
3	298	10.8	244	1 AEG1_MOUSE	Q03401 mus musculu
4	298	10.8	245	1 CRG3_HORSE	Q19010 equus cabal
5	297.5	10.8	245	1 CRG3_HUMAN	P54108 homo sapien
6	281	10.2	243	1 TPX1_HUMAN	P16562 homo sapien
7	269	9.7	312	1 YR81-CAEBL	Q09566 caenorhabdi
8	268.5	9.7	244	1 TPX1-CAVPO	Q60477 cavia porce
9	267.5	9.7	243	1 TPX1_MOUSE	P16563 mus musculu
10	257.5	9.3	242	1 HELO_HELHO	Q01055 heloderma h
11	256	9.2	227	1 VAS_VESPU	Q05110 vesputia vul
12	253	9.2	204	1 VAS_VESPU	P35783 vesputia fla
13	244	8.8	204	1 VAS_VESPE	P35785 vesputia pen
14	243.5	8.8	234	1 VAS_SOLIN	P35778 solenopsis
15	240	8.7	204	1 VAS_VESMC	P35760 vesputia mac
16	239.5	8.7	249	1 CRG1_HUMAN	P54107 homo sapien
17	238	8.6	202	1 VAS1_VESCR	P35781 vesputia ger
18	238	8.6	204	1 VAS_VESGE	P35784 vesputia ger
19	238	8.6	206	1 VAS_VESVI	P35787 vesputia vid
20	237	8.6	202	1 VAS2_VESCR	P35782 vesputia crabr
21	237	8.6	203	1 VAS_DOLAR	Q05108 dolichovesp
22	237	8.6	211	1 VAS_SOLRI	P35779 solenopsis
23	236.5	8.6	424	1 ASP_ANCCA	Q16377 ancylostoma
24	235	8.5	202	1 VAS_VESMA	P81657 vesputia manda
25	235	8.5	205	1 VAS_VESPO	P35786 vesputia equ
26	233.5	8.5	205	1 VAS_POUFU	P35780 polistes fu
27	232	8.4	173	1 PRL_MEDTR	Q40374 medicago tr
28	230	8.3	206	1 VAS_POIDO	P81656 polistes ex
29	222.5	8.1	205	1 VAS_POLEX	P35759 polistes ex
30	222.5	8.1	209	1 VAS3_POLAN	Q05109 polistes an
31	221.5	8.0	215	1 VAS3_DOLMA	P10737 dolichovesp
32	219.5	8.0	183	1 CRVP_TRIMU	P79845 trimeresuru
33	217	7.9	227	1 VAS2_DOLMA	P10736 dolichovesp

ALIGNMENTS

RESULT	1	ALIGNMENTS
GLIP_HUMAN	STANDARD;	PRT; 266 AA.
ID	GLIP_HUMAN	
AC	P48060: 015409;	
DT	01-FEB-1996 (Rel. 33, Created)	
DT	01-NOV-1997 (Rel. 35, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	Gloma pathogenesis-related protein (Glpr) (RTVP-1 protein).	
GN	GLPR OR RTVP1.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
OX	NCBI-TaxID-9606;	
RP	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=glial tumor;	
RX	MEDLINE=97128816; PubMed=8973356;	
RT	Rich T., Chen P., Furman F., Huynh N., Israel M.A.;	
RT	"RTVP-1, a novel human gene with sequence similarity to genes of	
RT	diverse species, is expressed in tumor cell lines of glial but not	
RT	neural origin.";	
RL	Gene 180:125-130(1996).	
RN	[2]	
RP	SEQUENCE OF 11-266 FROM N.A.	
RX	MEDLINE=9531646; PubMed=7607567;	
RT	Murphy E.V., Zhang Y., Zhu W., Biggs J.;	
RT	"The human glioma pathogenesis-related protein is structurally	
RT	related to plant pathogenesis-related proteins and its gene is	
RT	expressed specifically in brain tumors.";	
RL	Gene 159:131-135(1995).	
RN	[3]	
RP	STRUCTURE BY NMR.	
RX	MEDLINE=98151500; PubMed=9482873;	
RT	Szyperski T., Fernandez C., Mumenthaler C., Wuehrlich K.;	
RT	"Structure comparison of human glioma pathogenesis-related protein	
RT	Glpr and the plant pathogenesis-related protein P1a indicates a	
RT	functional link between the human immune system and a plant defense	
RT	system.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 95:2262-2266(1998).	
CC	-1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE HUMAN BRAIN TUMOR.	
CC	GLOBULINOMY MULTIFORM/ASTROCYTOMA, BUT NEITHER IN NORMAL FETAL	
CC	OR ADULT BRAIN TISSUE, NOR IN OTHER NERVOUS SYSTEM TUMORS.	
CC	-1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;	
CC	INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.	
CC	-----	
CC	THIS SWISS-PROT ENTRY IS COPYRIGHTED. IT IS PRODUCED THROUGH A COLLABORATION	
CC	BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -	
CC	THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS	
CC	USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY	
CC	MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL	
CC	ENTITIES REQUIRES A LICENSE AGREEMENT (SEE http://www.isb-sib.ch/announce/	
CC	OR SEND AN EMAIL TO license@sib-sib.ch).	
CC	-----	
DR	EMBL: X91911; CAA63005.1;	P35794 schizophyl1
DR	EMBL: U16307; AAA82731.1;	Q03402 mus musculu
DR	HSPF: P04284; ICFE.	P08299 nicotiana t
		P11670 nicotiana t
		Q04108 lycopersico
		Q00008 zea mays (m
		P07053 nicotiana t
		P09042 nicotiana t
		P35795 schizophyl1
		P35792 hordeum vul
		P35798 hordeum vul
		Q05968 hordeum vul

DR MIM: 602692; -
 DR InterPro: IPR001283; SCP.
 DR Pfam: PF00188; SCP, 1.
 DR PRINTS: PR00837; V5FXPLKE.
 DR SMART: SM00198; SCP, 1.
 DR PROSITE: PS01009; SCP_AGS_PRL_SC7_1; 1.
 DR PROSITE: PS01010; SCP_AGS_PRL_SC7_2; 1.
 FT CONFLICT 125 125 D->N (IN REF. 2).
 FT CONFLICT 209 266 RORDQKRYSVYVPGMPLIPNRYTSPLVNSVILLAS
 ITTIVOLKRYPLVLLD -> DSEKSNVTTMLYRLAHIS
 T (IN REF. 2)
 SQ SEQUENCE 266 AA; 30342 MW; C04378085FEB2B7A CRC64;

Query Match
 Best Local Similarity 12.9%; Score 355; DB 1; Length 266;
 Matches 82; Conservative 33; Mismatches 75; Indels 40; Gaps 12;

QY 25 LPNVLLEELSKYOHNSHVRRAIPRED-KEEILMLHKLKGQVQPOASNEEYMTWD 83
 D 5 LATIMWVSFVSNYSHF---ANILPDIENEDPDKCVRIHKKFSEVKEPTASDMLYMTWD 61
 QY 84 DELESAAMAASOCIWEHG---PTSL--LVSIGONLGAHGRYSPPGHYS---W 131
 D 62 PALAOIAKAMASNCOFSHNTRLRKPPHKLHPFTSLGENI---W-TGSVPLTFSVSAITNW 117
 QY 132 YDEKVDYTPYPSGCPNCPGRCGPMCTHTOTVMAVTKNGCAVNTCRKMTVGEWE 191
 D 118 YDEIQQDYDF-----KTRICKKVCGHYTOVMADSKYKVGCAVOFCRKSQGF-DALS 166
 QY 192 NAVFVONYSFKGN---WISAPYKNGRPGSCPPSYGSGCRNNLCYREE 238
 D 167 NGAFHICNYGPGGNYPTW---PYKRCATCSACPN--DKCDNLVCVNRQ 210

RESULT 2
 AEG_RAT STANDARD: PRT; 246 AA.

AC P12020;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, last annotation update)
 DE Sperm-coating glycoprotein precursor (SCP) (acidic epididymal glycoprotein) (Protein D) (Protein E) (Protein IV) (Sialoprotein)
 DE (32 kDa epididymal protein).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxId=10116;
 RP SEQUENCE FROM N.A.
 RC TISSUE-Epididymis;
 RC MEDLINE=87053955; PubMed=3780731;
 RA Brooks D.E., Means A.R., Wright E.J., Singh S.P., Tiver K.K.;
 RT "Molecular cloning of the cDNA for androgen-dependent sperm-coating glycoproteins secreted by the rat epididymis."
 RL Eur. J. Biochem. 161:13-18(1986).
 RN 12)
 RP SEQUENCE FROM N.A.
 RC MEDLINE=89039913; PubMed=2460753;
 RA Charrest N.J., Joseph D.R., Wilson E.M., French F.S.;
 RT "Molecular cloning of complementary deoxyribonucleic acid for an androgen-regulated epididymal protein: sequence homology with metalloproteins."
 RL Mol. Endocrinol. 2:999-1004(1988).
 CC -1- FUNCTION: THIS PROTEIN IS SUPPOSED TO HELP SPERMATOZOA UNDERGO FUNCTIONAL MATURATION WHILE THEY MOVE FROM THE TESTIS TO THE DUCTUS DEFERENS.
 CC -1- INDUCTION: By androgens.
 CC -1- MISCELLANEOUS: SCP IS AN ANDROGEN-DEPENDENT PROTEIN, WHICH IS SECRETED BY THE EPIDIDYMAL EPITHELIUM AND THEN BECOMES ASSOCIATED WITH THE SPERM SURFACE. TWO MAJOR VARIANT PROTEIN D AND E DIFFER FROM EACH OTHER BY THEIR CARBOHYDRATE SIDE CHAINS.

-1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1; INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: X04643; CA28304.1; -
 DR EMBL: M31173; BAB59716.1; -
 DR PIR: A24609; A24609.
 DR PIR: A40918; A40918.
 DR InterPro: IPR001283; SCP.
 DR Pfam: PF00188; SCP, 1.
 DR PRINTS: PR00837; V5FXPLKE.
 DR SMART: SM00198; SCP, 1.
 DR PROSITE: PS01009; SCP_AGS_PRL_SC7_1; 1.
 DR PROSITE: PS01010; SCP_AGS_PRL_SC7_2; 1.
 KW Sperm; glycoprotein; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 246
 FT MOD. RES 20 20
 FT CARBOHYD 32 32
 FT CARBOHYD 85 85
 FT CARBOHYD 147 147
 FT CARBOHYD 213 213
 SQ SEQUENCE 246 AA; 27847 MW; 585207C7CF7CE9D1 CRC64;

Query Match
 Best Local Similarity 10.9%; Score 301.5; DB 1; Length 246;
 Matches 81; Conservative 51; Mismatches 90; Indels 81; Gaps 15;

QY 11 LGILFLVCGSGYILPNTVLEELSKYOHN-ESISRRRAIPREDKEEILMLHKLKGQ 69
 D 5 LVLEFLAA-----VLP-PSLIQDTTDEMDRLDENTSTKLSV---OEEILINKHQLRRT 54
 QY 70 VQPOASNEEYMTWDELEKSAAMAASOCIWEHG---TSLVLSIGONL-----CAHGRY 121
 D 55 VSPSSDILLREVMHDVAVNAQKMANRCIYNHSPLOHRTTLKKGEMLFMANPANSVSV 114
 QY 122 RSPGFHVSQWYDEKVDYTPYPSGCPNCPGRCGPMCTHTOTVMAVTKNGCAVNTCR 181
 D 115 -----IQDWDESLDLYVFGG-----PKV-GYKVGHYTOVMWNSVTLVACVAECP 160
 QY 182 KMTWGEWENAVFVGNYSFKGNMIGE--APYKNGRPGSCPPSYGSGCRNNLCYREE 239
 D 161 DQPL-----KYFYVCHYCPGCGNYVGRVLSPTTEGRPCSCP-----GNCEDGLCTNSCE 209
 QY 240 YTPPETDENNEVETARIPENHVMVLOPRMRTYKRTKTSAAVYMQVYKCDTK--KKDC 298
 D 210 Y-----EDNT-----SNGCDLKKMYSCDDPLKEEC 235
 QY 299 KGS 301
 D 236 RAS 238

RESULT 3

AEG1_MOUSE STANDARD: PRT; 244 AA.

AC 003401;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, last annotation update)
 DE Sperm-coating glycoprotein I precursor (SCP 1) (acidic epididymal glycoprotein I) (Cysteine-rich secretory protein-1) (Crisp-1).
 GN AEG1 OR AEG-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Submandibular gland;
 RX MEDLINE=93246016; PubMed=1301383;
 RA Mizuki N., Kasahara M.;
 RT "Mouse submandibular glands express an androgen-regulated transcript
 encoding an acidic epididymal glycoprotein-like molecule.";
 RL Mol. Cell. Endocrinol. 89:25-32(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Submandibular gland;
 RX MEDLINE=93307144; PubMed=8319566;
 RA Haendler B., Kratzschmar J., Theuring F., Schleuning W.D.;
 RT "Transcripts for cysteine-rich secretory protein-1 (CRISP-1; DE/ABC)
 and the novel related CRISP-3 are expressed under androgen control in
 the mouse salivary gland";
 RL Endocrinology 133:192-198(1993).
 CC -1- FUNCTION: THIS PROTEIN IS SUPPOSED TO HELP SPERMATOZOA UNDERGO
 CC FUNCTIONAL MATURATION WHILE THEY MOVE FROM THE TESTIS TO THE
 CC DUCTUS DEFERENS.
 CC -1- SUBCELLULAR LOCATION: STORED IN SECRETORY GRANULES OF GRANULAR
 CC CONVOLUTED TUBULES CELLS.
 CC -1- TISSUE SPECIFICITY: MAINLY FOUND IN THE CAUDA EPIDIDYMIS WHERE IT
 CC IS SYNTHESIZED BY THE PRINCIPAL CELLS AND SECRETED INTO THE LUMEN.
 CC BURNS TO THE HEADS OF SPERMATOZOA. ALSO EXPRESSED IN THE
 CC SUBMANDIBULAR GLAND.
 CC -1- DEVELOPMENTAL STAGE: EXPONENTIAL INCREASE BETWEEN DAYS 25 AND 30
 CC AFTER BIRTH.
 CC -1- INDUCTION: By androgens.
 CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
 CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M92849; AAA37185.1; -
 DR EMBL: M05559; AAA37460.1; -
 DR PIR: A49202; A49202.
 DR MGI: MGI:102553; Aeg1.
 DR InterPro: IPR001283; SCP.
 DR Pfam: PF00188; SCP; 1.
 DR PRINTS: PRO0837; V5TPXLIKE.
 DR SMART: SM00198; SCP; 1.
 DR PROSITE: PS01009; SCP_AG5_PRL_SC7_1; 1.
 DR PROSITE: PS01010; SCP_AG5_PRL_SC7_2; 1.
 DR Sperm; Glycoprotein; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 244
 FT CARBOHYD 145 145
 FT SEQUENCE 244 AA; 27679 MW; D0DD0348F85781F CRC64;
 SO
 Query Match 10.8%; Score 298; DB 1; Length 244;
 Best Local Similarity 32.0%; Pred. No. 1,1e-16;
 Matches 77; Conservative 35; Mismatches 87; Indels 42; Gaps 11;
 OY 11 IGLFLVCGSGGYLTPNVLLELLSKYOHNSHVRRAIPREDKEELIMLNKIRGOV 70
 DB 5 LVLFELAA-----VLP-PSILOD-SSQENRLEKLTSTKMSV-----QEEIVSKHNOLERRV 53
 OY 71 QPQASNMETWTWDELKSAAMAASOCIEWHEGPTSLVS-----IGQNLG-----AHGRRV 122
 DB 54 SPSSGSDILKEMNYDAQVNAQOMADCKTSHSPILRTTLNRCGNLFMSSTYLAASS-- 111
 OY 123 SPGHVQSWYDEVKDYTPYPSPSCNPNWPCRGCPMCKHTOTIATVATNKIGCAVNTCKR 182
 DB 123 SPGHVQSWYDEVKDYTPYPSPSCNPNWPCRGCPMCKHTOTIATVATNKIGCAVNTCKR 182

DB 112 -----AIOGWNEYKDLTY-----DYGRKQPDSPVYG-----HYQVWNSTROVAGNAECRK 159
 OY 183 MTWGEWENAVFYVCNYSGKNWIGE--APYKNGRPCSCPSYSGSCNNLCYREETY 240
 DB 160 NPL-----RYYYVCHYCPVGNYGRLYTPYRAGEPCASCPCPHCEDGLCTNCGHEDEX 212
 OY 241 T 241
 DB 213 T 213
 RESULT 4
 CR33 HORSE STANDARD; PRT: 245 AA.
 ID CR33 HORSE 019010;
 DI 15-JUL-1998 (Rel. 36, Created)
 DI 15-JUL-1998 (Rel. 36, Last sequence update)
 DI 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cysteine-rich secretory protein-3 precursor (CRISP-3).
 GN CRISP3.
 OS Equus caballus (Horse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
 OX NCBI_TaxID=9796;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Amputilla;
 RX MEDLINE=98422318; PubMed=9748582;
 RA Schambony A., Gentzel M., Wolfes H., Raida M., Neumann U.,
 RA Toepfer-Petersen E.;
 RT "Equine CRISP-3: primary structure and expression in the male genital
 RT tract";
 RL Biochim. Biophys. Acta 1387:206-216(1998).
 CC -1- SUBCELLULAR LOCATION: SECRETED; IN NEUTROPHILS. LOCALIZED IN
 CC SPECIFIC GRANULES (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE SALIVARY GLAND, IN THE
 CC AMPULLA AND THE SEMINAL VESICLE.
 CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
 CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AJ001400; CA04729.1; -
 DR HSBP; P04284; ICFE.
 DR InterPro: IPR001283; SCP.
 DR Pfam: PF00188; SCP; 1.
 DR PRINTS: PRO0837; V5TPXLIKE.
 DR SMART: SM00198; SCP; 1.
 DR PROSITE: PS01009; SCP_AG5_PRL_SC7_1; 1.
 DR PROSITE: PS01010; SCP_AG5_PRL_SC7_2; 1.
 DR Signal; Multigene family.
 FT SIGNAL 1 22
 FT CHAIN 23 245
 FT SEQUENCE 245 AA; 27308 MW; 8934AE87FE402BA22 CRC64;
 SO
 Query Match 10.8%; Score 298; DB 1; Length 245;
 Best Local Similarity 27.9%; Pred. No. 1,1e-16;
 Matches 77; Conservative 35; Mismatches 72; Indels 92; Gaps 13;
 OY 54 EKEKEELIMLNKIRGOVQPOASNMETWTWDELKSAAMAASOCIEWHEGPTSLVS-----LLVS 109
 DB 36 EVQKELVKNHNDLRLRYVPLASNMILKQMDSTATNAQNMANKLLOSKAEDRAVGTMK 95
 OY 110 IQQNTL-----GAHWGRYSPGFHVQSWYDEVKDYTPYPSPSCNPNWPCRGCPMCKHTOTI 164
 DB 96 CENELFMSSIFPMSWD-----AIONWHDVDFY-----GVGPTPAVAVG-----HTQ 141

```

FT      VARIANT      134          134          A -> S (IN DBSNP:18643112).
SQ      SEQUENCE     245 AA: 27630 MW: 84DD79CB7AE9E5F9 CRC64;
                                         /FTID=VAR_011719.

Query Match
Best Local Similarity 31.0%; Score 297.5; DB 1; Length 245;
Matches 75; Conservative 29; Mismatches 91; Indels 47; Gaps 10;

QY      13 LFLVCGSGOGLLPNTVLLEELSTKYOHNSHSVRRAI---PREDEETLMLHNRIGQ 69
        |||||
        7 LFLVAG-----LLPSFRANDKDPATFALTPTOTQVOARETVRNHNELRRA 52
        |
QY      70 VOPOASNMETMYMDDELESAAMASQCWEH-GPTSLIVSI--GNQL-----GAHWGRY 121
        | | | | |
        53 VSPFRANMLKMEWNKEAANAOKMAQNCYRHSPNDRMTSLKGENLYMSSASSMSQ- 111
        | | | | |
QY      122 RSPGFHVOSWYDEVKDYTPYPSECPMPGCRSCGMCHHTQIYMATNKIGCAVNTCR 181
        |||||
        Db      112 -----AIQSMEFYENDFE-----GVGPRTPNNAVGC-----HYIQVVYSYLVCGCNAACP 158
        |
QY      182 KMTVGGEVENAVPYVCANSPKGNWIGE--APYKNRGPCSECPSPYGSGSCRNNLCYREET 239
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      159 NOKVL-----KYYVCOCYPAGNMANRLVPYEGGAPCASCPPDNODDLCTINGCYEDL 212
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      240 YT 241
        :
Db      213 YS 214

RESULT      6
            TPX1_HUMAN
ID           TPX1_HUMAN              STANDARD:             PRT:    243 AA.
AC           P16562;
DT           01-AUG-1990 (Rel. 15, Created)
DIT          01-AUG-1990 (Rel. 15, Last sequence update)
DI           16-OCT-2001 (Rel. 40, Last annotation update)
DE           Testis-specific protein TPX-1 precursor (Cysteine-rich secretory
protein-2) (Crisp-2).
OS           TPX1.
GS           Homo sapiens (Human).
OC           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX           NCBI_TaxID=9606;
RN           [1]
RP           SEQUENCE FROM N.A.
RC           TISSUE=Testis.;
RX           MEDLINE=90129048; PubMed=2613236;
RA           Kasahara M., Gultknecht J., Brew K., Spurr N., Goodfellow P.N.;
RT           Cloning and mapping of a testis-specific gene with sequence
RL           similarity to a sperm-coating glycoprotein gene.*;
RL           Genomics 5:527-534(1989).
[2]
RP           SEQUENCE FROM N.A.
RC           TISSUE=Testis.;
RX           MEDLINE=96270732; PubMed=8665901;
RA           Kraetzschmar J., Haendler B., Eberspaecher U., Roostermann D.,
RT           Donner P., Schleuning W.-D.;
RT           "The human cysteine-rich secretory protein (CRISP) family. Primary
structure and tissue distribution of CRISP-1, CRISP-2 and CRISP-3.";
RL           Eur. J. Biochem. 236:827-836(1996).
CC           -1 SUBCELLULAR LOCATION: Secreted (probable).
CC           -1 TISSUE SPECIFICITY: TESTIS AND EPIDIDYMIS.
CC           -1 STIMULABILITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
INSECTS AG3/AG5; FUNGI SCT/SC14 AND PLANTS PR-1.

-----
Cc      This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
Cc      The European Bioinformatics Institute. There are no restrictions on its
Cc      use by non-profit institutions as long as its content is in no way
Cc      modified and this statement is not removed. Usage by and for commercial
Cc      entities requires a license agreement (See http://www.ebi.ac.uk/copyright/)

```

```

CC -----
DR EMBL: M25532; AAA6120.1;
DR EMBL: X95239; CAA64526.1;
DR PIR: B33329; B33329.
DR MIM: 187430;
DR InterPro: IPR001283; SCP.
DR Pfam: PF00837; V5TPX1KE.
DR PRINTS: SM00198; SCP.
DR SMART: SM00198; SCP.
DR PROSITE: PS01009; SCP_AG5_PRL_SCT_1;
DR PROSITE: PS01010; SCP_AG5_PRL_SCT_2;
DR Testis: Signal; Multigene family.
KW SIGNAL
FT CHAIN 1 21 POTENTIAL.
FT 22 243 TESTIS-SPECIFIC PROTEIN TPR-1.
SQ SEQUENCE 243 AA; 27259 MW; C5FE698C449CFAD9 CRC64;

```

```

Query Match 10.2%; Score 281; DB 1; Length 243;
Best Local Similarity 26.3%; Pred. No. 2.4e-15;
Matches 83; Conservative 33; Mismatches 83; Indels 116; Gaps 14;

```

```

QY 24 LIPNVTLLEELISKYOHNSHSVRRAIPREDKE-----ETLMLHNLKRG 68
DB 3 LIPVPLVLYLPL-----SLPAEGKDPATFALLTLOLOVORELYNKNELRK 49
QY 69 QVOPQASNMETWTWDELEKSAAMASOCIEHNG-PTSLVLS--IGONLGAHMGRRSP- 124
DB 50 AVSPASNMMLKMEWSREVTTNAORMANKCTLOHSDPEDKRTSGENL-----YMSD 103
QY 125 -----GFVHVSQYDEKDYTPYPSECPNCPGSCPMCTHTYQIYMATNKGCAVNTC 180
DB 104 PTSSSAIQSMDYDILDEYV-----GVGPKSPNAVVG---HYQLVWYSTYQGGCIATC 155
QY 181 RKMTVWGEWENAVYFCNCSPPKGNWIG--EAPYKNGRCPSECPSPSYGSCRNMLCYREE 238
DB 156 PNOD-----SLKYYVCQYCPAGNNMNRKRTFYQOSTPCAGCP-----DCDRLGLC----- 201
QY 239 TTPKPEIDENNEVETAPIPENHVLQPRVMPRTKRTSAVYMYQVRYCPT----- 292
DB 202 -----TNSCOYQDILLSNCDLSLKNFTAG 222
QY 293 -----KKDKCKGKSTC 303
DB 223 CEHELKECK-ATC 236

```

```

RESULT 7
ID YR81_CAEEL STANDARD: PRT: 312 AA.
AC 009566;

```

```

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 35.0 kDa protein F48B8.1 in chromosome III.

```

```

GN F48B8.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodermidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Kirsten J.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

or send an email to license@isb-sib.ch.

```

CC -----
DR EMBL: U23514; AAC46538.1;
DR HSSP: P04284; ICFE.
DR WormRep: F48B8.1; CE01953.
DR InterPro: IPR001283; SCP.
DR Pfam: PF00188; SCP.
DR PRINTS: PR00837; V5TPX1KE.
DR SMART: SM00198; SCP.
DR PROSITE: PS01009; SCP_AG5_PRL_SCT_1; FALSE NEG.
DR PROSITE: PS01010; SCP_AG5_PRL_SCT_2; FALSE NEG.
KW Hypothetical protein.
FT DOMAIN 265 268 POLY-THR.
FT 299 303 POLY-GLU.
SQ SEQUENCE 312 AA; 35054 MW; AEEC7BF25E26288 CRC64;

```

```

Query Match 9.7%; Score 269; DB 1; Length 312;
Best Local Similarity 28.4%; Pred. No. 2.9e-14;
Matches 83; Conservative 36; Mismatches 93; Indels 80; Gaps 17;

```

```

QY 21 QGVLNVTLLEL-ILSKYOH-NE-----SHSRVRAIPREDKEILMLHNLKRGV 70
DB 55 RGFFPSHFGSDGLSRSEHPEYIAKKTITHEHNRVRRMP----- 96
QY 71 QPQASNMETWTWDELEKSAAMASOCIEHNGPTSLVLSIGONLGAHMGRRSPGFHYOS 130
DB 97 -----ASDNMMLYMSDELAAASORHADTCDFRHSRGR--INVENIMA--APYNSDAISI 149
QY 131 WYDEKDYTPYPSECPNCPGSCPMCTHTYQIYMATNKGCAVNTCKRT- 184
DB 150 WFNELV-----NPRCGCHNAVKHC-----CGHYVQYVMAKTLVCGGSRREDVVG 195
QY 185 VMGEWENAVYFCNCSPPKGNWIG-----GE-----APYKNGRCPSECPSPSYGSGSCR 230
DB 196 VMGRGHRNV--FVCHYNDGNTVYFARGOLYMPAFMTASGNGK-CSCCPAN-APACY 251
QY 231 NNLCTREETYPKPEIDENNEVETAPIPENHVLQPRVMPRTKRTSAVYMYQVRYCPT 282
DB 252 QGLCYMPKNTAEAPTITTB--STTSTTTEE-----PTTCEPDEPEAGADN 296

```

```

RESULT 8
ID TPX1_CAVPO STANDARD: PRT: 244 AA.
AC 060477;

```

```

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Testis-specific protein TPX-1 precursor (Autoantigen 1) (25 kDa
DE acrosomal autoantigen) (AA1).
GN TPX1.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriocognathi; Cavidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HARTLEY; TISSUE=Testis;
RX MEDLINE=96354287; Pubmed=9115720;
RA Foster J.A., Gerton G.L.;
RT "Autoantigen 1 of the guinea pig sperm acrosome is the homologue of
RT mouse tpX-1 and human TPX1 and is a member of the cysteine-rich
RT secretory protein (CRISP) family."
RL Mol. Reprod. Dev. 44:221-229(1996).

```

```

RN [2]
RP SEQUENCE OF 22-41. Pubmed=3282555;
RX MEDLINE=88193219; Pubmed=3282555;
RA Hardy D.M., Huang T.T.F., Jr., Driscoll W.J., Tung K.S.K., Wild G.C.;
RT "Purification and characterization of the primary acrosomal
RT autoantigen of guinea pig epididymal spermatozoa."
RL Biol. Reprod. 38:423-437(1988).
CC -1- TISSUE SPECIFICITY: TESTIS.

```

CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL; U35712; AAC52616.1; -
DR HSSP; P04284; 1CFE.
DR InterPro; IPR001283; SCP.
DR Pfam; PF00188; SCP, 1.
DR PRINTS; PR00837; V5TPXLIKE.
DR SMART; SM00198; SCP, 1.
DR PROSITE; PS01009; SCP_AGS_PRL_SC7_1; 1.
DR PROSITE; PS01010; SCP_AGS_PRL_SC7_2; 1.
DR Testis; signal.
FT SIGNAL
FT CHAIN
SQ SEQUENCE 244 AA; 27248 MW; 58DBDE6ECEE16A12 CRC64;

Query Match
Best Local Similarity 9.7%; Score 268.5; DB 1; Length 244;
Matches 74; Conservative 29; Mismatches 87; Indels 67; Gaps 12;

56 KEELMLHNLKRGQVOPASNMETWDELEKSAAMASOCIMENG-----PTSLVSIQ 110
38 QREIYNKNEHRLRSVNPFGSDILKEMSIQATNAQKANKCIIHSSKDKKINIGCE 97
111 GQNL-----GAHGRYRSPGFHVQSWYDEKYDTYYPSECPNCPERCSPCHTYQI 164
96 GENLYMSDPSWSD-----AIOSEFDESQDFTFEGVSKSHN-----AVQCHYQ 140
165 IWATNKNKIGCAVNTCKRKTWGEVWENAVYVCNYSKGN--WIGEPYKNGRSCSCP 222
141 LVMYSYLVGCCIAVCPNOD-----SLKYVYCYCPAGNNVYTKNPKYKOGITCASC 194
223 PEGGSCRNMLCYREBYTPKPEFDEMNEVEFAPLPEENHWLQRPVWPKPKKTSVN 282
195 ---CHCENGICLTNCEY-----EDLLNCESL-----KNTACE 225
283 YMTQVVRCDTKMKDKCK 299
226 HQLLYEKC--KATCRCE 240

RESULT 9
TPX1_MOUSE
ID TPX1_MOUSE STANDARD; PRT: 243 AA.

AC P16563;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Testis-specific protein TPX-1 precursor.
GN TPX1 OR TPX-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=90129048; PubMed=2613236;
RA Kasahara M., Gutknecht D., Brew K., Spurr N., Goodfellow P.N.;
RT Cloning and mapping of a testis-specific gene with sequence
RL similarity to a sperm-coating glycoprotein gene.
CC -!- SUBCELLULAR LOCATION: Secreted (Probable).
CC -!- TISSUE SPECIFICITY: Testis.

CC -!- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL; M25533; AAA40472.1; -
DR PIR; A33329; A33329.
DR HSSP; P04284; 1CFE.
DR MGD; MGI:98815; TPX1.
DR InterPro; IPR001283; SCP.
DR Pfam; PF00188; SCP, 1.
DR PRINTS; PR00837; V5TPXLIKE.
DR SMART; SM00198; SCP, 1.
DR PROSITE; PS01009; SCP_AGS_PRL_SC7_1; 1.
DR PROSITE; PS01010; SCP_AGS_PRL_SC7_2; 1.
DR Testis; signal.
FT SIGNAL
FT CHAIN
SQ SEQUENCE 243 AA; 27605 MW; 6E707E569ACAA244 CRC64;

Query Match
Best Local Similarity 9.7%; Score 267.5; DB 1; Length 243;
Matches 63; Conservative 28; Mismatches 62; Indels 37; Gaps 9;

56 KEELMLHNLKRGQVOPASNMETWDELEKSAAMASOCIMENG-----PTSLVSIQ 112
38 QREIYNKNEHRLRSVNPFGSDILKEMSIQATNAQKANKCIIHSSKDKKINIGCE 97
113 NLGAHWGRYRSP-----GFHVQSWYDEKYDTYYPSECPNCPERCSPCHTYQI 167
98 NL-----YMTDPTLMSYVQSWYDEKYDTYYPSECPNCPERCSPCHTYQI 142
168 ATKIKICAVNTCKRKTWGEVWENAVYVCNYSKGN--WIGEPYKNGRSCSCP 224
143 YSSBKICGGLAVCPNOD-----DNLKFFVCHYCPAGNNVYTKNPKYKOGITCASC 193
225 YGSGCRNMLC 234
194 --NNCENGIC 201

RESULT 10
HELO_HELHO
ID HELO_HELHO STANDARD; PRT: 242 AA.

AC Q91055;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Heloderma precursor (HLTX).
OS Heloderma horridum horridum (Mexican beaded lizard).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Anguilliformia; Helodermatidae;
OC Heloderma.
OX NCBI_TaxID=8552;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC TISSUE=Salivary gland.
RX MEDLINE=95375162; PubMed=7647234;
RA Morrisette J., Kraitschmar J., Heandler B., El-Hayek R.,
RA Mochoa-Morales J., Martin B.M., Patel J.R., Moss R.L.,
RT Schleuning W.-D., Coronado R., Possant L.D.;
RT Primary structure and properties of helodermin, a peptide toxin
RL that blocks ryanodine receptors.
CC Biophys. J. 68:2280-2288(1995).
CC SEQUENCE OF 20-39, AND CHARACTERIZATION.

RC TISSUE-Venom;
 RA PubMed-1693019;
 RX Mochca-Morales J., Martin B.M., Possani L.D.;
 RT "Isolation and characterization of hebeltherine, a novel toxin from
 Heloderma horridum horridum (Mexican beaded lizard) venom.";
 RL Toxicon 28:299-309(1990).
 CC -1- FUNCTION: Toxin that blocks ryanodine receptors. It is toxic to
 mice; it causes lethargy, partial paralysis of rear limbs and
 lowering of body temperature.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: Produced by the venomous gland
 CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
 CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: U13619; AAC59730.1; -
 DR HSSP: P04284; ICEE.
 DR Interpro: IPR001283; SCP.
 DR Pfam: PF00188; SCP. 1.
 DR PRINTS: PR00837; V5TPXLIKE.
 DR SMART: SM00198; SCP. 1.
 DR PROSITE: PS01009; SCP_AG5_PRL_SC7_1; 1.
 DR PROSITE: PS01010; SCP_AG5_PRL_SC7_2; FALSE_NEG.
 DR Calcium channel inhibitor; Neurotoxin; Toxin; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 242
 FT SEQUENCE 242 AA; 27493 MW; 0E183FC2F925DF3C CRC64;
 SQ
 Query Match 9.3%; Score 257.5; DB 1; Length 242;
 Best Local Similarity 31.5%; Pred. No. 1.8e-13;
 Matches 64; Conservative 28; Mismatches 80; Indels 31; Gaps 6;
 QY 54 EDEELIMLNKRGQVQPOASNMETWMDLEKSAAMASOCIMRHGP---TSLVLS 109
 DB 34 DQGTETDKNNLRIVEPASPNNLKTWNSKTAQNSANCTEHTSKEERTIDGVE 93
 QY 110 IGNLGAHMGVRSPPGHVQSWDEVYDVTYPPESECNPMCPERCSCP-----MCTHYTQ 164
 DB 94 CGENLPESSAPY-TWSTAIQNMEDERKTYRNFY-----GPTAQNVNIGHYQ 139
 QY 165 IYVATTNRKIGCAVNTCRKMTWGEWENAVYFCNYSKPNWIGE--APYKNGRPSCEP 222
 DB 140 VVMYRSYEICCAIAYCPDQPTY-----KYQVQYCPGGINRSKTYPSIGPGCDP 193
 QY 223 PSYGGSCNNLCYRETYTPKE 245
 DB 194 DACDNGLCTNPKQNDYNNCPD 216
 RESULT 11
 VAS_VESVU STANDARD: PRT; 227 AA.
 AC 005110; 090B91;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Venom allergen 5 precursor (Antigen 5) (AG5) (Allergen Ves v 5) (Ves v
 V).
 OS Vespaula vulgaris (Yellow jacket) (Vasp).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
 OC Vespidae; Vespinae; Vespinae; Vespa.
 OC NCBI_TaxID=7454;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE-Venom;
 RA MEDLINE-93203603; PubMed-8454859;
 RX Lu G., Villalba M., Coscia M.R., Hoffman D.R., King T.P.;
 RT "Sequence analysis and antigenic cross-reactivity of a venom
 RT allergen, antigen 5, from hornets, wasps, and yellow jackets.";
 RL J. Immunol. 150:2823-2830(1993).
 CC [2]
 CC SEQUENCE OF 24-227 FROM N.A.
 RA Suck R., Hagen S., Fiebig H.;
 RT "Molecular cloning of a genomic sequence from the venom allergen
 RT antigen 5 from Vespa vulgaris.";
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBI databases.
 CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
 CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: M98858; AAA30333.1; -
 DR HSSP: A0238849; CAB42887.1; -
 DR Interpro: IPR001283; SCP.
 DR Pfam: PF00188; SCP. 1.
 DR PRINTS: PR00837; V5TPXLIKE.
 DR SMART: SM00198; SCP. 1.
 DR PROSITE: PS01009; SCP_AG5_PRL_SC7_1; 1.
 DR PROSITE: PS01010; SCP_AG5_PRL_SC7_2; 1.
 DR Venom; Allergen; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 227
 FT DISULFID 27 40
 FT DISULFID 31 124
 FT DISULFID 49 117
 FT DISULFID 193 210
 FT CONFLICT 109 109
 FT CONFLICT 118 118
 FT CONFLICT 173 173
 FT CONFLICT 219 219
 FT SEQUENCE 227 AA; 25798 MW; 99E9813740A66F55 CRC64;
 SQ
 Query Match 9.3%; Score 256; DB 1; Length 227;
 Best Local Similarity 30.5%; Pred. No. 2.2e-13;
 Matches 71; Conservative 33; Mismatches 73; Indels 56; Gaps 12;
 QY 1 MSCVIGVITPLGLFLVCGSOGYLLPNVTLLLELLSKYQHNESHRSVRAIPREDKEIL 60
 DB 29 IKCLKGV-----HTAC-KYGSILKPNCG--NKVVSV-----GLTKQKQDIL 68
 QY 61 MLHNKLGQV-----QPOASNMETWMDLEKSAAMASOCIMRHGP---TSLVLS 108
 DB 69 KEHNDFRQKIAKGLDTRGNPNPGPPAKNMKNLVNDELAYVAQVANOCQGHOTCRDVA 128
 QY 109 --SIGONL--GAHWGRFRSPGFHVQSWDEVYDVTYPPESECNPMCPERCSCP---MCT 160
 DB 129 KYQVQVAVLGTSGRAKDDPYKLYKMEDEVKY-----NP--KKKSGNDFLKTG 178
 QY 161 IYVATTNRKIGCAVNTCRKMTWGEWENAVYFCNYSKPNWIGEAPYK 213
 DB 179 HYTGVMVANTKEVGG-----SIKYIQKMKH-KLYVNCYSGNFMNEELIYQ 225
 RESULT 12
 VAS_VESFL STANDARD: PRT; 204 AA.
 AC 005110; 090B91;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)

```

DE 01-NOV-1997 (Rel. 35, Last annotation update)
DT Venom allergen 5 (Antigen 5) (AG5) (Allergen Yes f 5) (Yes f V).
OS Vesputia flavopilosa (yellow jacket) (wasp).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
OC Vespidae; Vespidae; Vespinae; Vespula.
OX NCBI_TaxID=30211;
RN [1]
RP SEQUENCE.
RC Tissue-Venom;
RX MEDLINE=94044316; PubMed=8227862;
RA Hoffman D.R.;
RT "Allergens in Hymenoptera venom. xxv: The amino acid sequences of
RT antigen 5 molecules and the structural basis of antigenic cross-
RT reactivity."
RL J. Allergy Clin. Immunol. 92:707-716(1993).
CC - FUNCTION: MAY HAVE AN ANCESTRAL FUNCTION IN THE PROMOTION OF
CC OVUM FERTILIZATION BY SPERM.
CC - SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1,
CC INSCST AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
CC PIR; B44522; B44522.
CC PIR; A44583; A44583.
CC HSSP; P04284; 1CPE.
CC InterPro; IPR001283; SCP.
CC Pfam; PF00188; SCP; 1.
CC PRINTS; PR00837; V5PEPLIKE.
CC SMART; SM00198; SCP; 1.
CC PROSITE; PS01009; SCP_AG5_PRL_SC7_1; 1.
CC PROSITE; PS01010; SCP_AG5_PRL_SC7_2; 1.
CC Venom; Allergen.
FT DISULFID 4 17 BY SIMILARITY.
FT DISULFID 8 101 BY SIMILARITY.
FT FT 26 94 BY SIMILARITY.
FT DISULFID 170 187 BY SIMILARITY.
SQ SEQUENCE 204 AA; 23274 MW; 7667232356AB2FC5 CRC64;

Query Match 9.2%; Score 253; DB 1; Length 204;
Best Local Similarity 30.5%; Pred. No. 3.3e-13;
Matches 71; Conservative 32; NonMatches 74; Indels 56; Gaps 12

OY 1 MSCVGLGVPLGLFLVLCGSGGLYLPNTVLLBELLSKYOHNSHSVRYRAIIRDEKEELT 60
DB 6 INCLAGV-----HTAC-KYSLKPNCG--NKVVVSY-----GLTKQEKODILT 45
OY 61 MHNKLRGV-----QPOASNEMTWMDTELEKSAANAASQCIEWHGPTSLIV 108
DB 46 KEHNDFRQKIAGLETRGNPGRPAKKMKNLVWDELALAYVAQVAMNQGYGHDICRDLA 105
OY 109 --SIGNL---GAMHGRYKSPGFVHQSWDEYKQDVTTPRPSNCNWCPRKSGP--MCT 160
DB 106 KYQVQNAVLSTSTAKKYDDPVKLYKMWDEVDKY-----NP--KKKFGNNFLKTG 155
OY 161 HTYQVMTATNTKIGCAVNTCRKMTFWGEWENAVFYVCNYSKGMWIGAPAK 213
DB 156 HYTQVMTATNTKRVGCG-----SKRTIQEKMHKH-LYVCNYSKGMFNQDELTYQ 202

RESULT 13
ID VAS_VESPE STANDARD; PRT; 204 AA.
AC P33785;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Venom allergen 5 (Antigen 5) (AG5) (Allergen Yes p 5) (Yes p V).
OS Vesputia pensylvanica (Western yellow jacket) (Wasp).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
OC Vespidae; Vespidae; Vespinae; Vespula.
OX NCBI_TaxID=30213;
RN [1]
RP SEQUENCE.

```

```

RC TISSUE-Venom:
RX MEDLINE=94044316; PubMed=8227862;
RA Hoffman D.R.;
RT "Allergens in Hymenoptera venom. XXV: The amino acid sequences of
RL antigen 5 molecules and the structural basis of antigenic cross-
RL reactivity."
CC J. Allergy Clin. Immunol. 92:707-716(1993).
CC -1 SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCF/TPX1;
CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
DR PIR; C44522; C44522.
DR PIR; C44583; C44583.
DR HSSP; P04284; 1CFE.
DR InterPro; IPR001283; SCP.
DR Pfam; PF00188; SCP; 1.
DR PRINTS; SM00198; V5TPELKE.
DR SMART; SM00198; SCP; 1.
DR PROSITE; PS01009; SCP_AG3_PRL_SC7_1; 1.
DR PROSITE; PS01010; SCP_AG5_PRL_SC7_2; 1.
KW Venom; Allergen.
FT DISULFID 4 17 BY SIMILARITY.
FT DISULFID 8 101 BY SIMILARITY.
FT DISULFID 26 94 BY SIMILARITY.
FT DISULFID 170 187 BY SIMILARITY.
SQ SEQUENCE 204 AA; 23317 MW; 85BD971066C7D7C8 CRC64;

Query Match 8.8%; Score 244; DB 1; Length 204;
Best Local Similarity 30.4%; Pred No.1.7e-12;
Matches 73; Conservative 29; Mismatches 68; Indels 70; Gaps 13.

QY 1 MSCVAGVPIPLGLFLVCGSGYGLPNTVLTLELLSKYQHNESHKVRRAIPREDKEIL 60
DB 6 IKLKGQV-----HTAC-KYGSLEKPCG--NKIVYS-----GLTKERKQDIL 45
QY 61 MLHNKLRGV-----QPQASNNEYTWQDDELEKSAANAQSCIMHEGPTSLIV 108
DB 46 KEHNDFRQKIANGLETGRGNPGPPPKKNNKILVMDELAAYVAQWANAQCYGHDCRCVA 105
QY 109 S--ISQNT---GAHWGRYKSPGFHVQSWYDEKDYTPYP---SECNPWCERCSPGC 159
DB 106 KYPGVGNVALTGSTADKYNDPVKLVKMEDEVKDYN--PKKFEENN-----FKKI 154
QY 160 THTYQIVATTKKKICAVATTCRKNVWGEV-----WENAVYFVCNYSKGNWIGADPYK 213
DB 155 GHYTWVWANTKEIGC-----GSIKYIOWEMHKH-YLVNYSGPSGNFGNEELYQ 202

RESULT 14
VA3_SOLIN STANDARD: PRT: 234 AA.
ID VA3_SOLIN STANDARD: PRT: 234 AA.
AC P35778; O16135; Q9TWZ2;
01-JUN-1994 (Rel. 29, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
DE Venom allergen itii precursor (Allergen Sol 1 3) (Sol 1 III).
DS Solenopsis invicta (Red imported fire ant).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
OC Formicidae; Myrmicinae; Solenopsis.
OX NCBI_TaxID=13686;
RN RN
RP SEQUENCE FROM N.A.
RC TISSUE-Venom:
RA Hoffman D.R.; Farrar D., Schmidt M., McConnell T.J.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 23-234 FROM N.A.
RX MEDLINE=96051059; PubMed=8588684;
RA Hoffman D.R.;
RT "Fire ant venom allergy."
RL Allergy 50:535-544(1995).
RN [3]
RP SEQUENCE OF 23-234.

```

CC TISSUE-Venom: PubMed=8423273;
 RX MEDLINE=93139387; PubMed=8423273;
 RA Hoffman D.R.;
 RT "Allergens in Hymenoptera venom XXIV: the amino acid sequences of
 RT Imported fire ant venom allergens Sol I II, Sol I III, and Sol I
 RT IV";
 RL J. Allergy Clin. Immunol. 91:71-78(1993).
 RN [4]
 RP PARTIAL SEQUENCE OF 23-45.
 RC TISSUE-Venom:
 RA MEDLINE=90285439; PubMed=2355158;
 RA Hoffman D.R., Smith A.M., Schmidt M., Moffitt J.E., Guralnick M.;
 RT "Allergens in Hymenoptera venom. XXII. Comparison of venoms from two
 RT species of imported fire ants, *Solenopsis invicta* and *richteri*.";
 RL J. Allergy Clin. Immunol. 85:988-996(1990).
 CC -1- DISEASE: THE MOST COMMON CAUSE OF INSECT VENOM ALLERGY IN THE
 CC SOUTHEASTERN UNITED STATES IS THE IMPORTED FIRE ANT.
 CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCF/TPX1;
 CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF012919; AAB65434.1; -
 DR PIR: B37330; B37330.
 DR PIR: C44582; C44582.
 DR HSSP: P04284; 1CPE.
 DR InterPro: IPR001283; SCP.
 DR Pfam: PF00188; SCP; 1.
 DR PRINTS: PR00837; V5TPXLIKE.
 DR SMART: SM00198; SCP; 1.
 DR PROSITE: PS01009; SCP_AGS_PRL_SCF_1; 1.
 DR PROSITE: PS01010; SCP_AGS_PRL_SCF_2; 1.
 DR Venom; Allergen; Signal.
 KW SIGNAL 1 22
 FT CHAIN 23 234
 FT DISULFID 26 41
 FT DISULFID 31 125
 FT DISULFID 52 118
 FT DISULFID 198 216
 FT DISULFID 183 183
 FT CONFLICT 194 194
 FT CONFLICT 199 199
 FT SEQUENCE 234 AA; 26351 MW; 539F510B59941D83 CRC64;
 QY Query Match 8.8%; Score 243.5; DB 1; Length 234;
 Best Local Similarity 35.0%; Pred. No. 2.2e-12;
 Matches 63; Conservative 23; Mismatches 61; Indels 33; Gaps 7;
 QY 55 DKEELIMLNKRGV-----QPOASNMEYMTWDELLEKSAAMASQCTWEHG 102
 DB 64 EKDATVNMHNELRQVAGSGKRGNGPQPAVKMKNLWDELATIAORMANOCFHEHD 123
 QY 103 PLSL--VSIGONLGA--HMGRRS--PGFHVSQWDEVKDY---TYPPSCNPMPCPR 153
 DB 124 ACRNVERFAVGONINATSSCKNKSTPNEMILLWYNEVDFDNKRISFSDNIIIM--- 180
 QY 154 CSGPACTHTQIWMATNKGCAVNTCRKMTWGEWENAVYFVCNYSKGMWIGEPAPYK 213
 DB 181 ----KVGHVTOIWMATNKGCAVNTCRKMTWGEWENAVYFVCNYSKGMWIGEPAPYK 213
 RESULT 15
 ID VAS_VESMC STANDARD: PRT; 204 AA.
 AC P35760;
 DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Venom allergen 5 (Antigen 5) (AG5) (Allergen Ves m 5) (Ves m V).
 OS Vespa maculifrons (Eastern yellow jacket) (Wasp).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
 OC Vespoidea; Vespidae; Vespinae; Vespula.
 OX NCBI_Taxid=7453;
 RN [1]
 RP TISSUE-Venom:
 RC MEDLINE=93203603; PubMed=8454859;
 RA Lu G., Villalba M., Coscia M.R., Hoffman D.R., King T.P.;
 RT "Sequence analysis and antigenic cross-reactivity of a venom
 RT allergen, antigen 5, from hornets, wasps, and yellow jackets.";
 RL J. Immunol. 150:2823-2830(1993).
 CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCF/TPX1;
 CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
 CC -----
 DR PIR: B37329; B37329.
 DR HSSP: P04284; 1CPE.
 DR InterPro: IPR001283; SCP.
 DR Pfam: PF00188; SCP; 1.
 DR PRINTS: PR00837; V5TPXLIKE.
 DR SMART: SM00198; SCP; 1.
 DR PROSITE: PS01009; SCP_AGS_PRL_SCF_1; 1.
 DR PROSITE: PS01010; SCP_AGS_PRL_SCF_2; 1.
 DR Venom; Allergen.
 KW SIGNAL 4 17
 FT DISULFID 8 101
 FT DISULFID 26 94
 FT DISULFID 170 187
 FT SEQUENCE 204 AA; 23332 MW; 8B68A94C41390311 CRC64;
 QY Query Match 8.7%; Score 240; DB 1; Length 204;
 Best Local Similarity 29.6%; Pred. No. 3.6e-12;
 Matches 68; Conservative 27; Mismatches 85; Indels 50; Gaps 9;
 QY 1 MSCVLGVPIPLGLFVCGSQGYLPPNVLLEELLSKYQHNSHSHVRAIPREDKEIL 60
 DB 6 IKLKAGV-----HMAC-KYSLKPNK-----GNKKYVSYLTTQEKODIL 45
 QY 61 MLNKLKRGV-----QPOASNMEYMTWDELLEKSAAMASQCTWEHGPTSLV 108
 DB 46 KEHNDPROKILARGLETRGNGPQPAVKMKNLWSEDLAYIAQWMANOCQYGHDTCDVA 105
 QY 109 --SIGONL--GAMHGRYSRPGFHVSQWDEVKDYTPYPSSEONPMPCPRKSGPMCTHT 163
 DB 106 KYVGONVALTGSTAATYNDPVKLVKMWEDVADYDPKKKFSSENNFL-----KIGHY 158
 QY 164 QIWMATNKGCAVNTCRKMTWGEWENAVYFVCNYSKGMWIGEPAPYK 213
 DB 159 QMWMATNKGCG-----SIKYQEMWKKH-YLVCVYGPNGNQNDELIV 202
 Search completed: May 25, 2002, 14:39:57
 Job time: 662 sec

6

SEA ID NO: 2

AC NO: AAY41738, Database: A-Geneg-032802

File Copy

117 HMGTRAGGFFHVSQMYDEKDYTPYBSECNPCPCRGSGMCTHTYOTVATNTKICGA 176
 3 nmgvyrpffivgwydevkdytppscnmpcpcrgsgmcthtyotvattkicga 62
 177 VNTGKRTKGEWENAVYFCNYSFKGNWIGEPYKNGRSCPCPSYSGSGCRNNLCYR 236
 63 vntcrtkrtkgevenavayfcnyspknwlgapaykngrpscepsysgscrnlycr 122
 237 EETYPKRETDNENETAPIPEENHWLQPRVMPRTKRTSAVNYMTQVVRCDTKMKD 296
 123 eetyprkretdevenetapipeenhwlgprvmprtkrtksavnymtqvvrctdkmkd 182
 297 RCKGSTCNRYOCFACCLNHNKAKIFGSLFYESSSSICRAAIHYGIIDDKGGLVDITRNGV 356
 183 rckgstcnryocpactclnknkafgslfyesssicroaahygilddkgglvdltrngv 242
 357 PFFVYSEHGVOSTSKITPSSSEMYSKYVODLDCYTTVAOLCPFEKPAITHCPRIHCPAH 416
 243 pffvserhgvostskitpsssemyskvovodldcyyttvaolcpefkpaithcprihcapah 302
 417 GDEPSTYAPVFGTNTYATDSGICKTAVHAGVLSNMSGDVTVMVDDKRTYVGSIRNGV 476
 303 cdepytapyvfgntnyatdsckctavhagvlsnmsgdvtvmvddkrttyvgsirngv 362
 477 QSESLGTPRDGKAFRIFAVRQ 497
 363 qseslgtprdgkafrifavrq 383

RESULT 7
 AAB93979 standard; Protein; 371 AA.
 AAB93979:
 26-JUN-2001 (first entry)
 Human protein sequence SEQ ID NO:14059.
 Human; primer; detection; diagnosis; antitense therapy; gene therapy.
 Homo sapiens.
 EP1074617-A2.
 07-FEB-2001.
 28-JUL-2000; 2000EP-0116126.
 29-JUL-1999; 99JP-0248036.
 27-AUG-1999; 99JP-0300253.
 11-JAN-2000; 2000JP-0118776.
 02-MAY-2000; 2000JP-0183767.
 09-JUN-2000; 2000JP-0241899.
 (HELI-) HELIX RES INST.
 Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 Ishii S, Sugiyama T, Wakematsu A, Nagai K, Otsuki T;
 WPI; 2001-318749/34.
 Primer sets for synthesizing polynucleotides, particularly the 5602
 full-length cDNAs defined in the specification, and for the detection
 and/or diagnosis of the abnormality of the proteins encoded by the
 full-length cDNAs -
 Claim 8; SEQ ID 14059; 2537bp + CD ROM; English.
 The present invention describes primer sets for synthesizing 5602
 full-length cDNAs defined in the specification. Where a primer set
 comprises: (a) an oligo-dT primer and an oligonucleotide complementary

to the complementary strand of a polynucleotide which comprises one of
 the 5602 nucleotide sequences defined in the specification, where the
 oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 of an oligonucleotide comprising a sequence complementary to the
 complementary strand of a polynucleotide which comprises a 5'-end
 sequence and an oligonucleotide comprising a sequence complementary to a
 polynucleotide which comprises a 3'-end sequence where the
 oligonucleotide comprises at least 15 nucleotides and the combination of
 the 5'-end sequence/3'-end sequence is selected from those defined in
 the specification. The primer sets can be used in antisense therapy and
 in gene therapy. The primers are useful for synthesizing polynucleotides,
 particularly full-length cDNAs. The primers are also useful for the
 detection and/or diagnosis of the abnormality of the proteins encoded by
 the full-length cDNAs. The primers allow obtaining of the full-length
 cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 represent oligonucleotides, all of which are used in the exemplification
 of the present invention.

Sequence 371 AA:

Query Match 75.1%; Score/2071; DB 22; Length 371;
 Best Local Similarity 99.7%; Pred. No. 2.7e-169;
 Matches 369; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 MSCVIGVPIPLGLFLVCGSGYLIPVTLLEELSKYOHNEHSRYRRAIPREDKEIL 60
 1 mscvlgvpiplglflvcsqgyllpvtlleellskynheshrraripredkeell 60
 61 MHNKLGQVQPOASNNMYTMDLEKSAANAASQCIWEHGPSLSLVSGONTGAMGR 120
 61 mhnklrgvqpqasnmeytmdleksaanaasqciwehgpsllsvsgnlgamgr 120
 121 YRSPGFHVOSKMYDEKRYTPYBSECNPCPCRGSGMCTHTYOTVATNTKICAVNTC 180
 121 yrspgfhwskmydevkdytppscnmpcpcrgsgmcthtyotvattntkicavntc 180
 121 yrspgfhwskmydevkdytppscnmpcpcrgsgmcthtyotvattntkicavntc 180
 181 RKMTVGEWENAVYFCNYSFKGNWIGEPYKNGRSCPCPSYSGSGCRNNLCYREET 240
 181 rkmtvgevenavayfcnyspknwlgapaykngrpscepsysgscrnlycreet 240
 241 TPKEPDEMNEVETAPIPEENHWLQPRVMPRTKRTSAVNYMTQVVRCDTKMKDRCG 300
 241 tpkedemnevetapipeenhwlgprvmprtkrtksavnymtqvvrctdkmkdrckg 300
 301 STCNRYOCFACCLNHNKAKIFGSLFYESSSSICRAAIHYGIIDDKGGLVDITRNGV 360
 301 stcnryocpactclnknkafgslfyesssicroaahygilddkgglvdltrngv 360
 361 KSERHGVOSTSKITPSSSEMYSKYVODLDCYTTVAOLCPFEKPAITHCPRIHCPAH 416
 361 kserhgvostskitpsssemyskvovodldcyyttvaolcpefkpaithcprihcapah 416
 361 kserhgvostskitpsssemyskvovodldcyyttvaolcpefkpaithcprihcapah 416

RESULT 8

AAY41738 standard; Protein; 500 AA.

AAY41738:

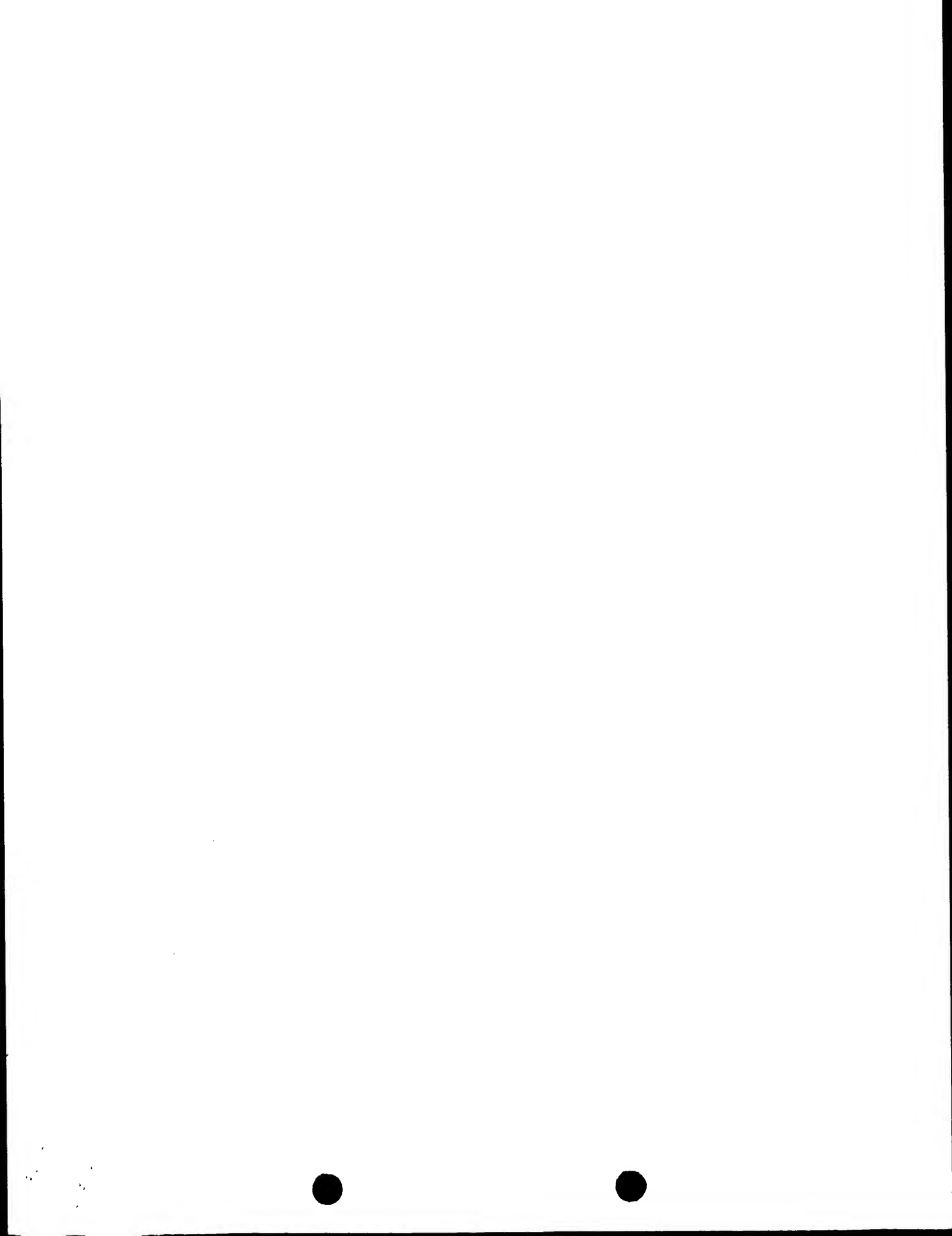
07-DEC-1999 (first entry)

Human PRO541 protein sequence.

Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
 probe; blood coagulation disorder; cancer; cellular adhesion disorder;
 secreted protein; transmembrane protein.

Homo sapiens.

MO9946281-A2.



SEA ID NO: 1
AC NO: AAD17766

Database: N-Geneseq-032802

421 cccaccagagagtgcaacccctggtgtccagagaggtgtcgggctatgtgcag 480
 426 cccctccagagagtgcaacccctggtgtccagagaggtgtcgggctatgtgcag 485
 481 cactacacacagatagtttgggcccacccaagaagatcgtgtgtcgtgcacacctgc 540
 486 cactacacacagatagtttgggcccacccaagaagatcgtgtgtcgtgcacacctgc 545
 541 cggagaagatgactgtctcgggagagagtttgggagacgcggttacttgcgtcaattat 600
 746 cggagaagatgactgtctcgggagagagtttgggagacgcggttacttgcgtcaattat 805
 601 tctccaaaggggaactgtgttggagagagagagagagagagagagagagagagagag 660
 806 tctccaaaggggaactgtgttggagagagagagagagagagagagagagagagagag 865
 661 tggccacccagctatgag 720
 866 tggccacccagctatgag 925
 721 actccaaacactgaaacgagagagagagagagagagagagagagagagagagagag 780
 926 actccaaacactgaaacgagagagagagagagagagagagagagagagagagagag 985
 781 aacatgttctgctccacacgagagagagagagagagagagagagagagagagagag 840
 986 aacatgttctgctccacacgagagagagagagagagagagagagagagagagagag 1045
 841 gtcaactacatgacccaagctgtcagatgtatgacatgaagatgaagagagagagag 900
 1046 gtcaactacatgacccaagctgtcagatgtatgacatgaagatgaagagagagagag 1105
 901 tccacgtgtacagagatgacagatgacagagagagagagagagagagagagagagag 960
 1106 tccacgtgtacagagatgacagatgacagagagagagagagagagagagagagagag 1165
 961 ggaagctgttctatgaaagctgtcgtatgacatgacagagagagagagagagagag 1020
 1166 ggaagctgttctatgaaagctgtcgtatgacatgacagagagagagagagagagag 1225
 1021 ctgagatgacag 1080
 1226 ctgagatgacag 1285
 1081 aagctctgag 1140
 1286 aagctctgag 1345
 1141 gtgtcaaaagtgaaagtgcaagatttgacgtcacaagcagctgtcgtgtgcag 1200
 1346 gtgtcaaaagtgaaagtgcaagatttgacgtcacaagcagctgtcgtgtgcag 1405
 1201 ttgtgaaagcagcaactactactcccaagaatcattgtccgacacactgtgaaagagaa 1260
 1406 ttgtgaaagcagcaactactactcccaagaatcattgtccgacacactgtgaaagagaa 1465
 1261 ccttcctactggtgtcgt 1320
 1466 ccttcctactggtgtcgt 1525
 1321 aagacagctgtgacgagcggt 1380
 1526 aagacagctgtgacgagcggt 1585
 1381 cccgtgtgataaaagagacactacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1440
 1586 cccgtgtgataaaagagacactacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1645
 1441 ctggt 1491
 1646 ctggt 1696

RESULT 7
 AAD17766
 ID AAD17766 standard; DNA; 2403 BP.
 AC AAD17766;
 XX 10-DEC-2001 (first entry)
 DE Human novel trypsin inhibitor-like protein, NOV-4d encoding DNA.
 KW Human: NOV-X protein; KIAA1233-like protein; STE20-like protein; tumour;
 KW trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
 KW immunological disorder; neurodegenerative disorder; Alzheimer's disease;
 KW Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
 KW human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
 KW cytosolic; nontropic; anti-fertility; cancer; NOV-4d protein; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 206..1702
 FT /tag= a
 FT /product= "Human novel trypsin inhibitor-like protein,
 FT NOV-4d"
 PN MO200162928-A2.
 PD 30-AUG-2001.
 PE 26-FEB-2001; 2001WO-US06151.
 PF 25-FEB-2000; 2000US-0184951.
 PR 28-FEB-2000; 2000US-0185548.
 PR 01-MAR-2000; 2000US-0185967.
 PR 18-APR-2000; 2000US-0197723.
 PR 27-APR-2000; 2000US-0199957.
 PR 23-FEB-2001; 2001US-0789390.
 PA (CURA-) CURAGEN CORP.
 PI Vernet CAM, Fernandes E, Shinkets RA, Macdougall J, Spaderna SK;
 DR WPI: 2001-582051/65.
 DR P-PSDB: AAEI7766.
 XX New isolated KIAA1233-like, STE20-like, or trypsin inhibitor-like
 PT polypeptide for diagnosing and treating pathological disorders, such as
 PT Parkinson's disease and for use in pharmacogenomics -
 PS Claim 9; Page 87-88; 189pp; English.
 XX The invention relates to novel human polypeptides referred as NOV-X
 CC and their corresponding nucleic acid sequences. NOV-X collectively
 CC include NOV-1, NOV-2a and NOV-2b which are novel KIAA1233-like
 CC polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel
 CC STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e
 CC which are novel trypsin inhibitor-like polypeptides. NOV-X is used to
 CC identify a potential therapeutic agent that can modulate its activity
 CC and can be used for treating a pathology related to aberrant expression
 CC or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
 CC used to determine the presence or predisposition to a disease associated
 CC with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
 CC to treat or prevent a pathology associated with NOV-X. The pathological
 CC states that can be treated or prevented are haematopoietic, cancer,
 CC immunological, tumour, neurodegenerative (e.g. Alzheimer's and
 CC Parkinson's disease), human immunodeficiency virus (HIV) illness and
 CC fertility disorders. NOV-X DNA and its DNA are used in pharmacogenomics for
 CC predictive medicine. NOV-X DNA is used in gene therapy. The present
 CC sequence is a DNA encoding human novel trypsin inhibitor-like protein,
 CC NOV-4d.
 XX Sequence 2403 BP; 558 A; 646 C; 685 G; 514 T; 0 other;

Oy	1018	atctcgagatgcgaaggagggcctgttgataataccacgaaacgggaaggtccccttttc	1077
Dd	1226	atctcgatctgcacaaggaggccctggctggataataccaccggaacgaggtcccccttcc	1285
Oy	1078	gtgaagtctgcagaacacagggcgttcgaatccctccagcaaatataaacctccaactcatc	1137
Dd	1286	gtgaagctctgcagagacagggcgttcgaatccctccagcaaatataaacctccaactcatc	1345
Oy	1138	atggtgtcaaaaagtgaagaatgcagagatttgacctgtcacacagccgttgtcagctgtgc	1197
Dd	1346	atggtgtcaaaaagtgaagaatgcagagatttgacctgtcacacagccgttgtcagctgtgc	1405
Oy	1198	cggtttgcgaagaacgacgacacacttgcaccaagaatccatctgttcgcgcgacactgcgaagac	1257
Dd	1406	cggtttgcgaagaacgacgacacacttgcaccaagaatccatctgttcgcgcgacactgcgaagac	1465
Oy	1258	gaaccttctaacttggctcccgctgttttggaaaccaacatctatgcagatatcctaagaatc	1317
Dd	1466	gaaccttctaacttggctcccgctgttttggaaaccaacatctatgcagatatcctaagaatc	1525
Oy	1318	tgcagaagaactctgtgcacgcggcgaggtcaatcaagaacgaagatgggggttgacgtgcagctg	1377
Dd	1526	tgcagaagaacgcgtgtgcacgcggcgaggtcaatcaagaacgaagatgggggttgacgtgcagctg	1585
Oy	1378	atgcccggtgcataaaaaagaagacttaacttggctcgcgtccaggaatgatgtcaatctgaa	1437
Dd	1586	atgcccggtgcataaaaaagaagacttaacttggctcgcgtccaggaatgatgtcaatctgaa	1645
Oy	1438	agccctgggagactcctcgggatlgtgaaaggccttcggatcctttgtctgcagcgag	1491
Dd	1646	agccctgggagactcctcgggatlgtgaaaggccttcggatcctttgtctgcagcgag	1699
<hr/>			
RESULT 8			
XX	AD17767	AD17767 standard; DNA; 2412 BP.	
XX	AC	AD17767;	
XX	DT	10-DEC-2001 (first entry)	
XX	DE	Human novel trypsin inhibitor-like protein, NOV-4e encoding DNA.	
KW	XX	Human: NOV-X protein; KIAA1233-like protein; SFE20-like protein; tumour	
KW	KM	trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;	
KW	KM	immunological disorder; neurodegenerative disorder; Alzheimer's disease;	
KW	KM	Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;	
KW	KM	human immunodeficiency virus; HIV; fertility disorder; neutropenic;	
KW	KM	cytostatic; neutropic; anti-fertility; cancer; NOV-4e protein; ds.	
XX	OS	Homo sapiens.	
XX	FH	Key Location/Qualifiers	
FT	CDS	206..1711	
FT		/*tag= a	
FT		/product= "Human novel trypsin inhibitor-like protein,	
FT		NOV-4e"	
FT		206..271	
FT	sig_peptide	/*tag= b	
FT		272..1708	
FT	mat_peptide	/*tag= c	
FT		/product= "Human mature novel trypsin inhibitor-like	
FT		protein, NOV-4e"	
XX	PN	WO200162928-A2.	
XX	PD	30-AUG-2001.	
XX	PE	26-FEB-2001; 2001IN US06151.	
XX	PR	25-FEB-2000; 2000US -0184951.	
XX	PR	28-FEB-2000; 2000US -0185548.	

SEA ID NO: 2
AC, NO: 99H0B8

Database: SP TREMBL-19

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:

May 25, 2002, 14:27:40 ; Search time 122.23 Seconds
(without alignments)
703.416 Million cell updates/sec

Title: US-09-667-380A-2

Sequence: 1 MSCVIGVPIGLFLVCGS.....SESLGTPRDGKAFRIFAVNG 497

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database :

1: SP.REMBL.19:*
2: SP.Archea:*
3: SP.Bacteria:*
4: SP.Fungi:*
5: SP.Human:*
6: SP.Invertebrate:*
7: SP.MHC:*
8: SP.Organelle:*
9: SP.Phage:*
10: SP.Plant:*
11: SP.Todent:*
12: SP.Virus:*
13: SP.Vertebrate:*
14: SP.Unclassified:*
15: SP.Virus:*
16: SP.Bacteriaph:*
17: SP.Archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2756	99.9	497	4 09H0B8	09H0B8 homo sapien
2	2071	75.1	371	4 096K61	096K61 homo sapien
3	1987.5	72.0	434	11 09D2R3	09D2R3 mus musculu
4	1640.5	59.5	500	4 09H336	09H336 homo sapien
5	1613	58.5	523	13 0985R5	0985R5 gallus galli
6	1536.5	55.7	507	11 099M06	099M06 mus musculu
7	968	35.1	188	11 0920U6	0920U6 rattus norv
8	758.5	27.5	258	13 0985T6	0985T6 gallus galli
9	740	26.8	258	11 099M07	099M07 mus musculu
10	739.5	26.8	258	4 043692	043692 homo sapien
11	646.5	23.4	253	4 09H3Y0	09H3Y0 homo sapien
12	639	23.2	120	4 096IR1	096IR1 homo sapien
13	357	12.9	415	5 044228	044228 halocynthia
14	355	12.9	266	4 0969K2	0969K2 homo sapien
15	349	12.6	489	11 09J556	09J556 mus musculu
16	349	12.6	489	11 09ET66	09ET66 mus musculu

17	323.5	11.7	236	11 09DAG6	09DAG6 mus musculu
18	319.5	11.6	255	11 09CWC1	09CWC1 mus musculu
19	301	10.9	244	11 091XA3	091XA3 mus musculu
20	294	10.7	332	11 09CQ35	09CQ35 mus musculu
21	293.5	10.6	203	4 09H108	09H108 homo sapien
22	292.5	10.6	203	4 09H106	09H106 homo sapien
23	286	10.4	380	5 09VFY2	09VFY2 drosophila
24	286	10.4	392	5 0960R5	0960R5 drosophila
25	273.5	9.9	301	5 0950F5	0950F5 caenorhabdi
26	273	9.9	220	5 016854	016854 onchocerca
27	273	9.9	243	11 09R1L4	09R1L4 rattus norv
28	268.5	9.7	243	11 088205	088205 rattus norv
29	254	9.2	334	6 09BE36	09BE36 macaca fasc
30	240	8.7	165	6 077720	077720 equus cabal
31	239	8.7	220	5 097149	097149 wuchereria
32	239	8.7	250	11 09D259	09D259 mus musculu
33	237	8.6	196	5 062507	062507 caenorhabdi
34	235	8.5	220	5 044932	044932 brugia mala
35	234.5	8.5	424	5 076744	076744 necator ame
36	234.5	8.5	424	5 044931	044931 onchocerca
37	234	8.5	220	5 044931	044931 onchocerca
38	234	8.5	249	6 09XSD3	09XSD3 macaca mula
39	233	8.4	176	10 093186	093186 aradidopsis
40	233	8.4	212	5 09BID5	09BID5 heterodera
41	232	8.4	212	5 096764	096764 caenorhabdi
42	231.5	8.4	208	5 018543	018543 caenorhabdi
43	231.5	8.4	217	6 077719	077719 equus cabal
44	230	8.3	190	10 095M05	095M05 aradidopsis
45	230	8.3	212	5 020609	020609 caenorhabdi

ALIGNMENTS

RESULT 1	PRELIMINARY	PRT	497 AA.
09H0B8			
ID 09H0B8			
AC 09H0B8			
DT 01-MAR-2001 (TREMBLrel. 16, Created)			
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE HYPOTHETICAL 55.9 KDA PROTEIN.			
GN DKEFZ434B044.			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX NCBI_TaxID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC TISSUE=TESTIS;			
RA Wambutt R., Heubner D., Mewes H.W., Gassenhuber J., Wleemann S.;			
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.			
DR EMBL; AL136861; CAB66795.1; ..			
DR HSP; P04284; ICFE.			
DR InterPro: IPR004043; LCOL.			
DR InterPro: IPR001283; SCP.			
DR Pfam: PF00188; SCP; 1.			
DR PRINTS: PR00837; V5TFPLIKE.			
DR SMART; SM00198; SCP; 1.			
DR PROSITE; PS01010; SCP_AGS_PRL_SCT_2; 1.			
KW Hypothetical protein.			
SQ SEQUENCE 497 AA; 55919 MW; 1E7ADDCAB37CA4ED CRC64;			

Query Match 99.9%; Score 2756; DB 4; Length 497;
Best Local Similarity 99.8%; Pred. No. 3.5e-259;

Matches 496; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSCVIGVPIGLFLVCGSOGYLLPNTLLLEILSKYOHNSHRVRAIPREDKEITL 60	09DAG6 mus musculu
DB	1	MSCVIGVPIGLFLVCGSOGYLLPNTLLLEILSKYOHNSHRVRAIPREDKEITL 60	09CQ35 mus musculu
QY	61	MLHNLKRGVOVPOASNMETWMDLEKSAANMASOCITWEGPISLVLSIGONICGAHWGR 120	09H0B8 homo sapien


```

Db 61 MLNKLKGQVOPQASNNEMTMDLEKSAAMASOCIMHEGPTSLVSGONLGAHGR 120
    |||
Qy 121 YRSPGFHVOSMYEDVDYTPYSECPNCPBCSGSMCHHTQIYATNKGCAVNTC 180
    |||
Db 121 YRSPGFHVOSMYEDVDYTPYSECPNCPBCSGSMCHHTQIYATNKGCAVNTC 180
    |||
Qy 181 RKMTVGEWENAVFYVCNTPSPKGNWIGEARPYKNGRCPSECPSPYSGSCNNLCYRETY 240
    |||
Db 181 RKMTVGEWENAVFYVCNTPSPKGNWIGEARPYKNGRCPSECPSPYSGSCNNLCYRETY 240
    |||
Qy 241 TPKPEIDENNEVEFAPLPEENHWMLOPRVMPKPKTSAVNTQYVRCDTKMKDRCKG 300
    |||
Db 241 TPKPEIDENNEVEFAPLPEENHWMLOPRVMPKPKTSAVNTQYVRCDTKMKDRCKG 300
    |||
Qy 301 STCNRYOCPAGCLNHRKAKIFGSLFYESSSSICRAAIHYGLIDKGLVDITRNGKVPFV 360
    |||
Db 301 STCNRYOCPAGCLNHRKAKIFGSLFYESSSSICRAAIHYGLIDKGLVDITRNGKVPFV 360
    |||
Qy 361 KSRHGVOSLSKRPSSSPWVKVQDDCYTVAQLCPFEKPAHCPRIHCPACKDE 420
    |||
Db 361 KSRHGVOSLSKRPSSSPWVKVQDDCYTVAQLCPFEKPAHCPRIHCPACKDE 420
    |||
Qy 421 PSYMAVFGTNIADTSSICKTAVHAGVISNESGGDVMPVDPKKTYYGSLNAGVQSES 480
    |||
Db 421 PSYMAVFGTNIADTSSICKTAVHAGVISNESGGDVMPVDPKKTYYGSLNAGVQSES 480
    |||
Qy 481 LGTPRDGKAFRIFAVRQ 497
    |||
Db 481 LGTPRDGKAFRIFAVRQ 497
    |||

```

RESULT 2

```

ID 096K61 PRELIMINARY: PRT: 371 AA.
AC 096K61;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CDNA FJ114489.FIS, CLONE YAMAMA100281, WEAKLY SIMILAR TO GLIOMA
DE PATHOGENESIS-RELATED PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MAMMARY GLAND;
RA Iisogai T., Oca F., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuhara Y., Sasaki N.;
RA "NGDO human cDNA sequencing project.";
RT Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL: AK027395; BAB55081.1;
SQ SEQUENCE 371 AA; 42807 MW; 8D23FBE14F53E85C CRC64;

```

Query Match

```

Best Local Similarity 75.1%; Score 2071; DB 4; Length 371;
Matches 369; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 MSCVIGVITPGLLEIVGSGYLLNNTLEELSTKQNHESRVRRAIPREDKEITL 60
Db 1 MSCVIGVITPGLLEIVGSGYLLNNTLEELSTKQNHESRVRRAIPREDKEITL 60
Qy 61 MLNKLKGQVOPQASNNEMTMDLEKSAAMASOCIMHEGPTSLVSGONLGAHGR 120
Db 61 MLNKLKGQVOPQASNNEMTMDLEKSAAMASOCIMHEGPTSLVSGONLGAHGR 120
Qy 121 YRSPGFHVOSMYEDVDYTPYSECPNCPBCSGSMCHHTQIYATNKGCAVNTC 180
Db 121 YRSPGFHVOSMYEDVDYTPYSECPNCPBCSGSMCHHTQIYATNKGCAVNTC 180
Qy 181 RKMTVGEWENAVFYVCNTPSPKGNWIGEARPYKNGRCPSECPSPYSGSCNNLCYRETY 240
Db 181 RKMTVGEWENAVFYVCNTPSPKGNWIGEARPYKNGRCPSECPSPYSGSCNNLCYRETY 240
Qy 241 TPKPEIDENNEVEFAPLPEENHWMLOPRVMPKPKTSAVNTQYVRCDTKMKDRCKG 300
Db 241 TPKPEIDENNEVEFAPLPEENHWMLOPRVMPKPKTSAVNTQYVRCDTKMKDRCKG 300
Qy 301 STCNRYOCPAGCLNHRKAKIFGSLFYESSSSICRAAIHYGLIDKGLVDITRNGKVPFV 360
Db 301 STCNRYOCPAGCLNHRKAKIFGSLFYESSSSICRAAIHYGLIDKGLVDITRNGKVPFV 360
Qy 361 KSRHGVOSLSKRPSSSPWVKVQDDCYTVAQLCPFEKPAHCPRIHCPACKDE 420
Db 361 KSRHGVOSLSKRPSSSPWVKVQDDCYTVAQLCPFEKPAHCPRIHCPACKDE 420
Qy 421 PSYMAVFGTNIADTSSICKTAVHAGVISNESGGDVMPVDPKKTYYGSLNAGVQSES 480
Db 421 PSYMAVFGTNIADTSSICKTAVHAGVISNESGGDVMPVDPKKTYYGSLNAGVQSES 480
Qy 481 LGTPRDGKAFRIFAVRQ 497
Db 481 LGTPRDGKAFRIFAVRQ 497

```

May

```

Qy 181 RKMTVGEWENAVFYVCNTPSPKGNWIGEARPYKNGRCPSECPSPYSGSCNNLCYRETY 240
    |||
Db 181 RKMTVGEWENAVFYVCNTPSPKGNWIGEARPYKNGRCPSECPSPYSGSCNNLCYRETY 240
    |||
Qy 241 TPKPEIDENNEVEFAPLPEENHWMLOPRVMPKPKTSAVNTQYVRCDTKMKDRCKG 300
    |||
Db 241 TPKPEIDENNEVEFAPLPEENHWMLOPRVMPKPKTSAVNTQYVRCDTKMKDRCKG 300
    |||
Qy 301 STCNRYOCPAGCLNHRKAKIFGSLFYESSSSICRAAIHYGLIDKGLVDITRNGKVPFV 360
    |||
Db 301 STCNRYOCPAGCLNHRKAKIFGSLFYESSSSICRAAIHYGLIDKGLVDITRNGKVPFV 360
    |||
Qy 361 KSRHGVOSLSKRPSSSPWVKVQDDCYTVAQLCPFEKPAHCPRIHCPACKDE 420
    |||
Db 361 KSRHGVOSLSKRPSSSPWVKVQDDCYTVAQLCPFEKPAHCPRIHCPACKDE 420
    |||

```

RESULT 3

```

ID 09D2R3 PRELIMINARY: PRT: 434 AA.
AC 09D2R3;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 1810049K24RIK PROTEIN.
GN 1810049K24RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX MEDLINE=21083660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana S.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fieischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Maesuo Y., Mikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli R., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Fujino M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Kamita M., Lee N.H.,
RA Gustincich S., Hill D., Holtmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez C., Seva T., Shibata Y., Storch K.-F.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shikata C., Shimada Y.,
RA Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kontsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RL EMBL: AK019034; BAB31519.1;
DR HSSP: P04284; ICPE.
DR MGD: MGI:1926142; 1810049K24RIK.
DR InterPro: IPR004043; LCCL.
DR InterPro: IPR001283; SCP.
DR Pfam: PF00188; SCP; 1.
DR PRINTS: PR00837; VSTPLIKE.
DR SMART: SM00198; SCP; 1.
DR PROSITE: PS01010; SCP_AGS_PRL_SCT_2.
SQ SEQUENCE 434 AA; 48593 MW; C25067E914647AC0 CRC64;

```

Query Match

```

Best Local Similarity 72.0%; Score 1987; DB 11; Length 434;
Matches 348; Conservative 37; Mismatches 49; Indels 3; Gaps 1;

```

```

Qy 61 MLNKLKGQVOPQASNNEMTMDLEKSAAMASOCIMHEGPTSLVSGONLGAHGR 120
    |||
Db 61 MLNKLKGQVOPQASNNEMTMDLEKSAAMASOCIMHEGPTSLVSGONLGAHGR 120
    |||
Qy 121 YRSPGFHVOSMYEDVDYTPYSECPNCPBCSGSMCHHTQIYATNKGCAVNTC 180
    |||
Db 121 YRSPGFHVOSMYEDVDYTPYSECPNCPBCSGSMCHHTQIYATNKGCAVNTC 180
    |||
Qy 181 RKMTVGEWENAVFYVCNTPSPKGNWIGEARPYKNGRCPSECPSPYSGSCNNLCYRETY 240
    |||
Db 181 RKMTVGEWENAVFYVCNTPSPKGNWIGEARPYKNGRCPSECPSPYSGSCNNLCYRETY 240
    |||
Qy 241 TPKPEIDENNEVEFAPLPEENHWMLOPRVMPKPKTSAVNTQYVRCDTKMKDRCKG 300
    |||
Db 241 TPKPEIDENNEVEFAPLPEENHWMLOPRVMPKPKTSAVNTQYVRCDTKMKDRCKG 300
    |||
Qy 301 STCNRYOCPAGCLNHRKAKIFGSLFYESSSSICRAAIHYGLIDKGLVDITRNGKVPFV 360
    |||
Db 301 STCNRYOCPAGCLNHRKAKIFGSLFYESSSSICRAAIHYGLIDKGLVDITRNGKVPFV 360
    |||
Qy 361 KSRHGVOSLSKRPSSSPWVKVQDDCYTVAQLCPFEKPAHCPRIHCPACKDE 420
    |||
Db 361 KSRHGVOSLSKRPSSSPWVKVQDDCYTVAQLCPFEKPAHCPRIHCPACKDE 420
    |||
Qy 421 PSYMAVFGTNIADTSSICKTAVHAGVISNESGGDVMPVDPKKTYYGSLNAGVQSES 480
    |||
Db 421 PSYMAVFGTNIADTSSICKTAVHAGVISNESGGDVMPVDPKKTYYGSLNAGVQSES 480
    |||
Qy 481 LGTPRDGKAFRIFAVRQ 497
    |||
Db 481 LGTPRDGKAFRIFAVRQ 497
    |||

```


XX Claim 2; Page 27-28; 29pp; English.

PS The present invention provides the protein and coding sequences of the
 CC novel human trypsin inhibitor like protein NHP. This shows homology to
 CC mammalian trypsin inhibitors. The sequences are useful in disease
 CC diagnosis and treatment, particularly of diseases associated with signal
 CC transduction. The present sequence is the NHP protein.

XX Sequence 497 AA:

Query Match 100.0%; Score 2759; DB 22; Length 497;
 Best Local Similarity 100.0%; Pred. No. 3.6e-228;
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSCVLGVPIPLGLFLVCSOGYLLPNVTLLEELSKYOHNSHSHVRAIRPREDKEEIL 60
 DB 1 mscvlgvpiplglflvcsogyllpnvtlleelskyohnseshvrrairpredkeell 60
 QY 61 MHNKLRGVVPOASNMEYMTWDELKSAAMASOCIEHGPTSLVSIQNTGAHNGR 120
 DB 61 mhnklrgvvpoasnmeymtwddelksaamaasociwebgptslvsiqnlghawgr 120
 QY 121 YRSPGFHVSQWTDVKKDYTPRPPSECNPCPCSGPMCTHTQIYMATNTKIGCAVNTC 180
 DB 121 yrspgfhwswtdvkkdytprppsecpncpcsgpmcthytqlvwaltnkigcavntc 180
 QY 181 RRMATWGEWENAVYFVCNYSFKGNMIGAPYKNGRPSCEPSPYSGSCNNLCYREETY 240
 DB 181 rrmatwgewenavyfvcnyspknmigeapyngrpscepspsysgscnnlcyreety 240
 QY 241 TPKEPTDENNETAPIPENHVMLOPRMPTPKTKTSANVTMVOVCDTRAKRCRG 300
 DB 241 tpkptdenneetapipenhvmloprrmptpktktsanvmtvmovcdtrmkrcrg 300
 QY 301 STCNRYOCPCAGCLNHRKAKIFGLFEYSSSSICRAAIHYGILDDKGLVDTNRGKVPFY 360
 DB 301 stcnryocpcagclnhkakifglfeyssssicraaihygilddkglvdtltnrgkvpfy 360
 QY 361 KSEBKGVSLSKTKRSSSMWVKVODIDCTTVAQLCPFEKPATHCPRINCAHCKDE 420
 DB 361 ksebrgvslsktkrsssmwvkvodidcttvaqlcpfekpahcprinhcackde 420
 QY 421 PSYMAVFCTNTYADTSSICTAVNAHGVTSNMSGVDVMPDPDKKTKVYGSILNGVQSSS 480
 DB 421 psymafvtntyadtssictavnahgvtsnmsgvdvmpdpdkkktvysilngvqss 480
 QY 481 LGTPRDGKAFFRIFAVRQ 497
 DB 481 lgtprdgkaffrifavrq 497

RESULT 2

AAE10616 standard; Protein: 497 AA.

AAE10616:

10-DEC-2001 (first entry)

Human novel trypsin inhibitor-like protein, NOV-4b.

Human; NOV-X protein; KIAA1233-like protein; STE20-like protein; tumour;
 trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
 immunological disorder; neurodegenerative disorder; Alzheimer's disease;
 Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
 human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
 cytoskeletal; motropic; anti-fertility; cancer; NOV-4b protein.

Homo sapiens.

Key Location/Qualifiers

FT Peptide 1..22
 FT /Label= Signal_peptide
 FT Protein 23..497
 FT /note= "Human mature novel trypsin inhibitor-like
 FT protein, NOV-4b"
 FT Domain 195..206
 FT /label= SCP_domain

WO200162928-A2.

30-AUG-2001.

26-FEB-2001; 2001WO-US06151.

25-FEB-2000; 2000US-0184951.

28-FEB-2000; 2000US-0185548.

01-MAR-2000; 2000US-0185967.

18-APR-2000; 2000US-0197723.

27-APR-2000; 2000US-0199957.

23-FEB-2001; 2001US-0789390.

(CDRA-) CUBAGEN CORP.

Vernet CAM, Fernandes E, Shinkets RA, Macdougall J, Spaderna SK;

WPI; 2001-582051/65.

N-PSDB; AAD17764.

New isolated KIAA1233-like, STE20-like, or trypsin inhibitor-like

polypeptide for diagnosing and treating pathological disorders, such as

Parinson's disease and for use in pharmacogenomics

Claim 1; Page 83; 189pp; English.

The invention relates to novel human polypeptides referred as NOV-X
 and their corresponding nucleic acid sequences. NOV-X collectively
 include NOV-1, NOV-2a and NOV-2b which are novel KIAA1233-like
 polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel
 STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e
 which are novel trypsin inhibitor-like polypeptides. NOV-X is used to
 identify a potential therapeutic agent that can modulate its activity
 and can be used for treating a pathology related to aberrant expression
 or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
 used to determine the presence or predisposition to a disease associated
 with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
 to treat or prevent a pathology associated with NOV-X. The pathological
 states that can be treated or prevented are haematopoietic, cancer,
 immunological, tumour, neurodegenerative (e.g. Alzheimer's and
 Parkinson's disease), human immunodeficiency virus (HIV) illness and
 fertility disorders. NOV-X and its DNA are used in pharmacogenomics for
 predictive medicine. NOV-X DNA is used in gene therapy. The present
 sequence is human novel trypsin inhibitor-like protein, NOV-4b.

Sequence 497 AA:

Query Match 99.4%; Score 2743; DB 22; Length 497;
 Best Local Similarity 99.6%; Pred. No. 8.6e-227;
 Matches 495; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSCVLGVPIPLGLFLVCSOGYLLPNVTLLEELSKYOHNSHSHVRAIRPREDKEEIL 60
 DB 1 mscvlgvpiplglflvcsogyllpnvtlleelskyohnseshvrrairpredkeell 60
 QY 61 MHNKLRGVVPOASNMEYMTWDELKSAAMASOCIEHGPTSLVSIQNTGAHNGR 120
 DB 61 mhnklrgvvpoasnmeymtwddelksaamaasociwebgptslvsiqnlghawgr 120
 QY 121 YRSPGFHVSQWTDVKKDYTPRPPSECNPCPCSGPMCTHTQIYMATNTKIGCAVNTC 180
 DB 121 yrspgfhwswtdvkkdytprppsecpncpcsgpmcthytqlvwaltnkigcavntc 180
 QY 181 RRMATWGEWENAVYFVCNYSFKGNMIGAPYKNGRPSCEPSPYSGSCNNLCYREETY 240

SEA ID NO: 2
 AC NO: AAE10616
 Database: A-Geneseg-032802

Db 181 rkmrtvgevenavayfvcnyspknwlgaeaykngdrpcsecpysgscrmnlycreey 240
 Qy 241 TPKEPTEDEMEVEETAPRPEENHVMLOPRVMPRTKRTSANYMTQVVRCDTKMKDRCKG 300
 Db 241 tpkpetdememevelaprppeenhvlpqrvmpkrtksaavymtqvvrctdkmkdrckg 300
 Qy 301 STCRNRYOCPCAGCINHKAKRTGSLFESESSSICRAAIHYGLIDDKGGLVDITRNGKVPFF 360
 Db 301 stcrnryocpcagcclnhkakrtgslfyeasssicraahyglliddkgglvdlitngkvpff 360
 Qy 361 KSEHGVOSLSKTKPSSFWKVKVODLDCYTTVAOLCPFEKPAHPCRIHCPAHCKDE 420
 Db 361 ksehgvoslsktpssfwskvkvdldcyltvaqlcpfekpachpcrihpcakde 420
 Qy 421 PSYMAVPGTNIYADTSSICKTAHVAGVISNESGSDVDVMPDKKTKYVGSILRNGVQSES 480
 Db 421 psywavpgtlniyadtsisicktahvavlsnesgsvdvmppdkkkyvgsilrnyvqses 480
 Qy 481 LGTPRDGKAFFRFAVRQ 497
 Db 481 lgtprdgkaffrlfavrq 497

RESULT 3
 AAE10618
 ID AAE10618 standard; Protein: 498 AA.
 AC AAE10618;
 DT 10-DEC-2001 (first entry)

Human novel trypsin inhibitor-like protein, NOV-4d.
 Human novel trypsin inhibitor-like protein, NOV-4d.
 Human, NOV-X protein; KIAA1233-like protein; STE20-like protein; tumour;
 trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
 immunological disorder; neurodegenerative disorder; Alzheimer's disease;
 Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
 human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
 cytosolic; nontoxic; anti-fertility; cancer; NOV-4d protein.

Homo sapiens.
 Key Location/Qualifiers
 Domain 196..207
 /label= SCP_domain

MO200162928-A2.
 30-AUG-2001.
 26-FEB-2001; 2001WO-US06151.
 25-FEB-2000; 2000US-0184951.
 28-FEB-2000; 2000US-018548.
 01-MAR-2000; 2000US-018548.
 18-APR-2000; 2000US-0197123.
 27-APR-2000; 2000US-0199557.
 23-FEB-2001; 2001US-0789390.
 (CORA-) CURAGEN CORP.

Vernet CAM, Fernandes E, Shinkets RA, Macdougall J, Spaderna SK,
 WPI; 2001-582051/65.
 N-PSDB; AAD17766.

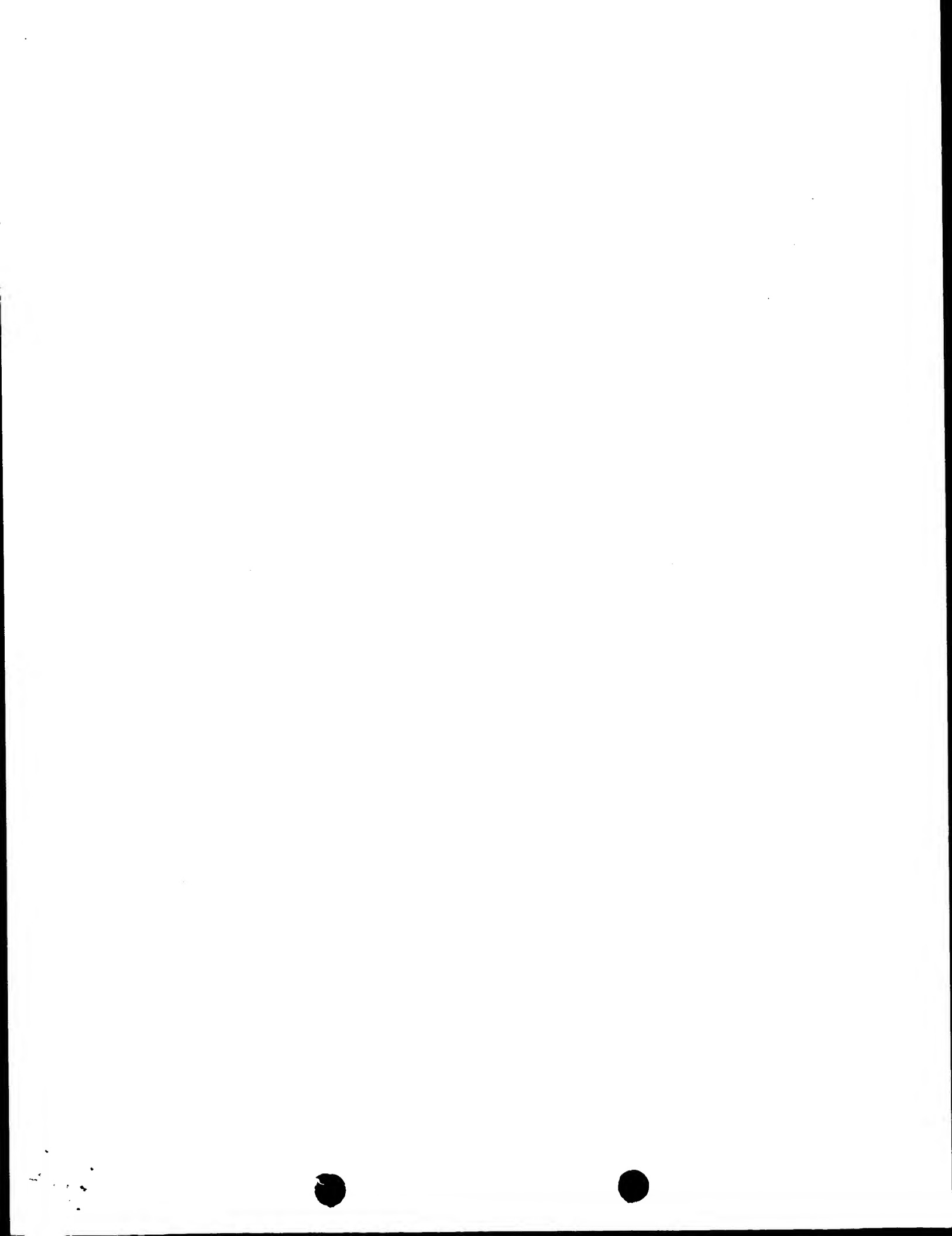
New isolated KIAA1233-like, STE20-like, or trypsin inhibitor-like
 polypeptide for diagnosing and treating pathological disorders, such as
 Parkinson's disease and for use in pharmacogenomics -
 Claim 1; Page 88; 189pp; English.

CC The invention relates to novel human polypeptides referred as NOV-X
 CC and their corresponding nucleic acid sequences. NOV-X collectively
 CC include NOV-1, NOV-2a and NOV-2b which are novel KIAA1233-like
 CC polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel
 CC STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e
 CC which are novel trypsin inhibitor-like polypeptides. NOV-X is used to
 CC identify a potential therapeutic agent that can modulate its activity
 CC and can be used for treating a pathology related to aberrant expression
 CC or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
 CC used to determine the presence or predisposition to a disease associated
 CC with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
 CC to treat or prevent a pathology associated with NOV-X. The pathological
 CC states that can be treated or prevented are haematopoietic, cancer,
 CC immunological, tumour, neurodegenerative (e.g. Alzheimer's and
 CC Parkinson's disease), human immunodeficiency virus (HIV) illness and
 CC fertility disorders. NOV-X and its DNA are used in pharmacogenomics for
 CC predictive medicine. NOV-X DNA is used in gene therapy. The present
 CC sequence is human novel trypsin inhibitor-like protein, NOV-4d.
 SQ Sequence 498 AA:

Query Match 99.2%; Score 2736.5; DB 22; Length 498;
 Best Local Similarity 99.6%; Pred. No. 3.1e-226;
 Matches 496; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1 MSCVLGVVPLILFLVYCSGGLPNTLLEELSKYQHNESHVRAIIPREDKEEIL 60
 Db 1 mscvlgvvplilflvyvcsqglpntlleelskqhyneshvrratpredkeell 60
 Qy 61 MLHNKLRGVQVPOASNMENMTWDELEKSAAMAASQCTHEPRTSLVSIQNTGAHMG- 119
 Db 61 mlhnklrgvgvqaemqemtwdeleksaasqctwehptslvsiqnlghwgr 120
 Qy 120 RRSRGFRVQSYDYNDYTPYSECNPCWPCERCSGPCCTHYTOIVATTKIGCAVNT 179
 Db 121 rrsrgfrvqsydyndytpypsecpwpcersgpccthytoivattkigcavnt 179
 Qy 180 CRKMTVGEVWENAVYFVCNYSKGNWIGEPYKNGRPSSECPSPYSGSCRNMLCYREBT 239
 Db 181 crkmtvgevenavayfvcnyspknwlgaeaykngdrpcsecpysgscrmnlycreey 240
 Qy 240 YTPKEPTEDEMEVEETAPRPEENHVMLOPRVMPRTKRTSANYMTQVVRCDTKMKDRCK 299
 Db 241 ytpkpetdememevelaprppeenhvlpqrvmpkrtksaavymtqvvrctdkmkdrck 300
 Qy 300 GSTCRNRYOCPCAGCINHKAKRTGSLFESESSSICRAAIHYGLIDDKGGLVDITRNGKVPFF 359
 Db 301 gstcrnryocpcagcclnhkakrtgslfyeasssicraahyglliddkgglvdlitngkvpff 360
 Qy 360 VKSEHGVOSLSKTKPSSFWKVKVODLDCYTTVAOLCPFEKPAHPCRIHCPAHCKD 419
 Db 361 vksehgvoslsktpssfwskvkvdldcyltvaqlcpfekpachpcrihpcakcd 420
 Qy 420 EPSTYMAVPGTNIYADTSSICKTAHVAGVISNESGSDVDVMPDKKTKYVGSILRNGVQSES 479
 Db 421 epsywavpgtlniyadtsisicktahvavlsnesgsvdvmppdkkkyvgsilrnyvqses 480
 Qy 480 SLGTPRDGKAFFRFAVRQ 497
 Db 481 slgtprdgkaffrlfavrq 498

RESULT 4
 AAE10619
 ID AAE10619 standard; Protein: 501 AA.
 AC AAE10619;
 DT 10-DEC-2001 (first entry)
 Human novel trypsin inhibitor-like protein, NOV-4e.



QY 1 atgagtgctgctctggtgtgtgtcattcccttggggctgtcttctctgtctcgagatcc 60
Db ATGAGTGTCTTCTTGAACAATATGGTCCGGTGGGGCTGGCTCCCTGGTGTGGGAGTA 145
QY 61 caaggtctactctctgcccacgctcaactctcttagaggagctctctagaataacagcac 120
Db AAGAGTGTCTTCTTCTTCCCAACACATGAGGCTGGAGAGGCTCTGACGATACAGCAC 205
QY 121 aacgagctcaactcccggtgtccgcaagacatcccccaggagagacaagagagatacttc 180
Db AACAGAGCCACATCTACGGGTGGCAGGGGCCAATCCCATCTGACAG-CCGTGAAATCTCTC 264
QY 181 atgctgcacaaagaactctcgggcgcaaggttcagctctcaggtctccacacattgag/aaabtg- 240
Db ATGCTGCACAAACACTCGCAGAGTCAAGGTATCCCTGCTCCAACTGAGATCACTG 324
QY 241 ~~atctctggatgttgaagaactgtagaagatctgtcgaagctgtggccatgcatgcatcgaggag~~ 300
Db AACTGGATGTAGAGACTGAGAGGTGGCAGCGCCGCTGGGCACAGAGGTGCTGTGGGAG 384
QY 301 caagcgccacacagctctgtctgtgtccatcgggcaaaacttgggcgtcactatggtggcagg 360
Db CATTGGCCCCGACGCTTCTGCTGTCATCGGGCAACCTGGCTGTGCTACATGGGGCAGG 444
QY 361 tatcgtctccgggtgtccatctgtgagctccgtgatctagaagagtgaaagagatacaactac 420
Db TAACGCTCTCCGTGGGTTCCAGGTGAGTGTGATGACGAGGTGAAGATTAACCTTAC 504
QY 421 cccctaccggagagtgcaacccctgtgtgtccagaagagtgctcgggtccatattgtgacg 480
Db CGTATACCCCCACGAGTGCACCCATGGTGGCCCCAGGCGCTCTCAGGAGGCATGTGCACC 564
QY 481 caactacacacagatagtttgtggccacacacacacacacacacacacacacacacacacac 540
Db CACTACACACAGATGTGTTGGGCCACACACACAAAGATGGCTGTGCTGTCACACCTGT 624
QY 541 cggagaagtactgtctgtgggagaaagtttgggagaaagcggtctacttctgtctgcaatat 600
Db CGGAGCATGACGTCTGGGGAGACATCTGGGAGAAACCCGTGATCTCTGTCTGCAATAT 684
QY 601 tctccaaggggaactgtgattgtgagaagccccctacaagaatgtgcggccctgtctgag 660
Db TCTCCAAAAGGGAAATCTGATTTGTGAGAGCTCCCTTCAACAAATGTGCGCCCTGCTTGAG 744
QY 661 tggccaccacagctatgtgagcagctgtcaggaaacaacttgyttacagagaagaacactac 720
Db TGTCTGTCACAGCTATGAGGTGGTGGCCGAACAACTTGTTGTTACGAGAAAGACACTTAC 804
QY 721 actccaaaaccttgaacggaacgagatgaaatgaaatgtggaadcggtccctcatctctgaaga 780
Db CATTAAGAAACC-GGAGTGGATGAGATGAAGAGGTGGAATCAACCCCTGCTCCGAGGAA 864
QY 781 aaacatttttggctcccaagcgaggtgatagtgagcccaacgcccagaagaacactctgcg 840
Db ACCCATTTCTGGGTTCAACCCAGGGTGTAAACCTTCAAAAGACCAAGAGACCCCAAGTC 924
QY 841 gtaactacatgacccaagtcgtcacaagtgtgacacacaagaatgaaagacaggtgtcgaagg 900
Db GTAAACTTATGACCAAGGTGTCCACTGTGACACAAAGTAAGTAAGTATGATATGATACCTTA 924

RESULT	15				
LOCUS	AX250094				
DEFINITION	AX250094	2664 bp	DNA	linear	PAT 28-SEP-2001
ACCESSION	AX250094	Sequence 93 from Patent WO0166747.			
VERSION	AX250094.1	GI:15864509			
KEYWORDS					
ORGANISM	human.				
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
FEATURES					
source					
	1..2664	Location/Qualifiers			
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	853 a	465 c	554 g	792 t	

Db	147	TGCTGTTATGGCTAGAGCAATTCACGACATGGGTTTCCCAATGGCACTTTATTTGAGA	206
QY	98	agctgctcagaanaattaccagacaagaagctcactcccggtcgc-----	144
Db	207	AACCTTTTGGAAAAATACATGGATGATGAGATGGTGATGGTGGATAGGCCAAACAGAGGA	266
QY	145	--agagccatccccaagagagagacaaggagatccctcactgctgcacacaagaacttcg	202
Db	267	AAAGGGCATCATCAGACAAATACATGCAGAGTATTTTGACCTTCATATTAATTACGAA	326
QY	263	agctcgtcgcagcgctgggcccagtcacagacatctgggagcccgccacagctcgtcg	322
Db	387	GATGTGCAGAACTCTGGGCTGAAAGTTGCTGTGGGAACATGGAACCTGCAGAGCTTCTTC	446
QY	323	tgctcatalcggcagaaccttgcgcgtcactccgggagagtagtcgctcgcgggttcctatg	382
Db	447	CATCAATTTGGACAGAAATTTGGAGCACACATGGGGAAATATAGGCCCCGACCTTTCATG	506
QY	383	tgacgtcttggtatgaagaggtgaagactcacactaccctaccggagcgatgcaacc	442
Db	507	TACATCTGTGTATGATGAAGATGAAAGTAAACACTTGTGCACCATATGAAACATGATGCAACC	566
QY	443	cctgtgtccagagaggtgtctcgggcctatgtgcacgcctcacacagatagtttgg	502
Db	567	CATATTGTGCATTCAGGTGTCTGGCCCTTATGTACACTTTATACACAGGTCGTGTGG	626
QY	503	ccacccccaacaatctcgtgtgtcgtgtgaacacctgcggagaagatgctctgtggag	562
Db	627	CAACTACTATACAGAAATCGGTTGTGCCATTAAATTTGTGTATATACATGAACATCTGGGGC	686
QY	563	aagttggaggaagcgcgttactctgtctcgaattattctccaaggggaactgattg	622
Db	687	AGATATGGCCCAAGCTGTACCTGCTGTGCATTAATCTCCCAAGGGAAATCGTGTGG	746
QY	623	gagaaagcccccatacaagaatgtccggccctcgtcctagtgcgccaaccagctatggaga	682
Db	747	GCCATGCCCCCTTACAAACAATGGCGGCCCGTGTCCTGTGCCACACTTAATTGTGGAGGG	806
QY	683	gtgcgaggaacaactgtgttaccgtagaagaaacctacatcccaaaacctgaacgagc	742
Db	807	GCTGTAGAGAAATTCGTGTCTACAAAAGAGGTCACACAGGTATTAATCCCTCGCAAG	866
QY	743	agatgaatgaggtggaacgcgtcccatctccgaagaacaatgttggctccacaaga	802
Db	867	AGGAACAAT---GAATATGAAACGACACACTTCACAAATCCATGACACCCATGCCGA	923
QY	803	gggtgtgagaccaccaagcccagaanaacctgtgcgttcaactacatgaccaagtcg	862
Db	924	CAAGATCAGATGATATATACAGAAATGAAGATCATTAAGCCACAGCAAAATGTCCAAATG	983
QY	863	tcaagtgtgaacccaagatggaaggcaagttgaagaaggttccacgtgtatacagatcaac	922
Db	984	TTTCTTGTGAAGTAAATTAAGATTAAGATCACTGCAAAAGGAACCAACCTGCATATAGTTCGAT	1043
QY	923	gcccaagcagctgcctgcaacccaagaagcgaagaatcttggaaatctgtgtcctatgaagct	982
Db	1044	GTCCTCTGCTGGCTGTTGGATGATGAAGCTAAAGTAAATTTGCGCATGTACATTTATGAAATGC	1103
QY	983	cgtctcagcatatgcgcgcgcgcacatccactacaggaatctctgatatcaagaagagcgtg	1042
Db	1104	AATCCAGCAATCTGTAGAGCTGCAATTCATTTAAGTATTAATAGACAAATATGCTGGCTGG	1163
QY	1043	tggatatccccaaggaacgggaaggttccctcttcgttgaagtctgagagacacggcgtgc	1102
Db	1164	TGATATCTACTGACAGGAAGAAAGCATTAATTTCATCATCAAGTCCAAATGAAATGTATTC	1223
QY	1103	agtcctcagcaaatatacaaaccttcacgctatcatcatgtgtgtcaaaagtgaagtgcag	1162
Db	1224	AAACAATTGGCAAAATATACGTCTGTATTTCTTCAACAGTCTCTTAAGATTAACATTGAG	1283

Tue May 28 16:10:03 2002

copy 2 (p. 2 + 3)

us-09-667-380a-2.rag

PD 9/24/1999 only 8 in ok (data), copy 6-8 Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2002, 13:22:04 ; Search time 76.94 Seconds

(without alignments)
717.490 Million cell updates/sec

Title: US-09-667-380A-2

Sequence: 1 MSCVUGGTYPLGLFLVCGS.....SGLCTPRDGRAFRFAVRQ 497

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A.Geneseq_032802.*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2759	100.0	497	22	AAAB74446
2	2743	99.4	497	22	AAAE10616
3	2736.5	99.2	498	22	AAAE10618
4	2721	98.6	501	22	AAAE10619
5	2595	94.1	477	22	AAAE10617
6	2140	77.6	383	22	AAAE10615
7	2071	75.1	371	22	AAAB39379
8	1640.5	59.5	500	20	AAAY41738
9	1640.5	59.5	500	21	AAAB44294
10	1640.5	59.5	500	22	AAU29058
11	1407.5	51.0	522	22	AAAM41693

ALIGNMENTS

12	1374	49.8	241	22	AAW24028	Rat EST encoded pr
13	1302	47.2	431	21	AAAB41816	Human OREX ORF1580
14	929	33.7	300	22	AAAB39907	Human polypeptide
15	915	33.2	300	22	AAU08687	Human ECTF7 polype
16	905.5	32.8	259	22	AAU20344	Human secreted pro
17	860.5	31.2	268	22	AAU20344	Human secreted pro
18	860.5	31.2	277	22	AAU20481	Human gene 1 encod
19	776	28.1	160	22	AAAE09683	Human protein sequ
20	730	26.5	198	16	AAAR79914	Tyrosin inhibitor
21	721	26.1	188	16	AAAR79915	Human gene 1 encod
22	661.5	24.0	233	22	AAAE09686	Human trypsin inh
23	661.5	24.0	233	22	AAAE09686	Human gene 1 encod
24	521	18.9	142	22	AAU20636	Human secreted pro
25	421	15.3	159	21	AAU19527	Human diagnostic a
26	421	15.3	159	21	AAU19527	Membrane-bound pro
27	421	15.3	159	22	AAU29143	Human PRO1156 (UNQ
28	355.5	12.9	348	22	AAAB5263	Novel human diagno
29	352	12.8	219	21	AAAB01400	Neuron-associated
30	351	12.7	266	22	AAAB1502	Neuron-associated
31	351	12.7	266	22	AAAB39716	Human polypeptide
32	351	12.7	463	20	AAAT13392	Amino acid sequenc
33	351	12.7	463	21	AAAB01373	Human PRO328 antit
34	351	12.7	463	21	AAAY95343	Human PRO328 polyp
35	351	12.7	463	22	AAAB88408	Human membrane or
36	351	12.7	463	22	AAAB88408	Human PRO328 prote
37	351	12.7	463	22	AAAB88408	Human angiotensin
38	351	12.7	463	22	AAAB88408	Gene 12 human secr
39	344.5	12.5	181	22	AAAB53088	Human cancer assoc
40	344	11.7	270	19	AAAB63115	A human pathogene
41	322	11.7	270	19	AAAB63115	Human EST encoded
42	297.5	10.8	245	22	AAAB23992	Human EST encoded
43	297.5	10.8	245	22	AAAB23992	Human PRO434 poly
44	295.5	10.7	242	22	AAU12302	Human full-length
45	295.5	10.7	258	22	AAE02211	

RESULT 1

AAAB74446 standard; Protein; 497 AA.

AAAB74446;

29-MAY-2001 (first entry)

Human protease-inhibitor like protein.

Human; protease-inhibitor like protein; NHP; novel human protein;

trypsin inhibitor; gene therapy; drug screening; disease diagnosis;

signal transduction.

Homo sapiens.

WO200121651-A2.

29-MAR-2001.

22-SEP-2000; 2000WO-US26048.

24-SEP-1999; 99US-0156101.

(LEXI-) LEXICON GENETICS INC.

Donoho G, Turner CA, Wattler F, Nehls M, Friedrich G;

Zambrowski B, Sands AT; WPI: 2001-266064/27.

N-ESDB; AAFT77686; AAFT77687. Novel isolated human protease inhibitor-like polynucleotide useful in therapeutic, diagnostic and pharmacogenic applications

SEA ID NO: 2
AE NO: AA10616
Database: A_GeneSeq-032802

XX
PS Claim 2; Page 27-28; 29pp; English.

AA The present invention provides the protein and coding sequences of the
CC novel human protease-inhibitor like protein NHP. This shows homology to
CC mammalian trypsin inhibitors. The sequences are useful in disease
CC diagnosis and treatment, particularly of diseases associated with signal
CC transduction. The present sequence is the NHP protein.

SQ Sequence 497 AA;

Query Match	100.0%	Score 2755	DB 228	Length 497
Best Local Similarity	100.0%	Pred. No. 3	6e-228	
Matches 497, Conservative	0	Mismatches	0	Indels 0; Gaps 0;

OY	1	MSCVJAGVJPLGJLFLVYCSQSJVLPRVNTLLEELSTKXONHNSRRBARJPRECKEEL	60
OY	1	meovjagvjlplgjlflvycsqjylprvntlleelstkyqjmeshsrvtrajpreckee	60
Db	1	meovjagvjlplgjlflvycsqjylprvntlleelstkyqjmeshsrvtrajpreckee	60
OY	61	MEHNKLRGQVOPOASNMJEWMTWDDLEKSAAMAASOCJWEHPTSLVSGONJGAHWGR	120
Db	61	mhnklrgvqpogpasmjeywtwddlekaasaawascjwehptslsvsgonjgahwgr	120
OY	121	YNSPGRNVSMTBENVDYTPRYSECPMPCEBSGPMCTHTOJYVMTNKGICAVNVC	180
Db	121	yspgrnvsmtbenvdyptrysecpmpcebsgpmcthtojyvmtnkgicavnc	180
OY	181	RKMTWGEVWENAVFYECNYSPKCNMTIGEARPYKNGRPCSECPSPYSGSCRNNLCYRETY	240
Db	181	rkmtwgevwenavyfecnyspkcnmtigearpykngrpcsecpspysgscrnnlcyreety	240
OY	241	TPRPEDDENNEVEFARJPREENHVMLOPRVMPRTPKTSANVMUQVYRCFKMDCRKG	300
Db	241	tprpedennevefarjpreenhvmloprvmprtpktsanvmuqvyrckmcdckrg	300
OY	301	STCNRYOCFAGCLNHNKAKITFGLSEYESSSSICRAIHGJLDDDKGGLVDTIRNGKVPFV	360
Db	301	stcnryocfagclnhnkakitfglseysesssicraihgjlddkgglvdtirngkvpfv	360
OY	361	KSERHGVQSJLSTKYRPSSEFMYSKXKVOOLDQYTTAOLCPEFKPHTHCRIHCPAHCXOE	420
Db	361	kserhgvqsjlstkyrpssefmyskxkvoldqyttaolcpefkphthcricphacxoe	420
OY	421	PSYPAPEFGENTYADTSSICTTAHAGVJSNESGGDVDMPYDKKRTYVGSJLNRGQSES	480
Db	421	psypapefgentyadtssicttaahagvjnesgddvdmpydkkrtysjlnrgqses	480
OY	481	LGTPRODKARIRIFAVRQ 497	
Db	481	lgtprogkairifavrq 497	

RESULT	2
AAE10616	

AAC10616;

DT 10-DEC-2001 (first entry)

DE Human novel trypsin inhibitor-like protein, NOV-4b.

Human; NOV-X protein; KIM1233-like protein; SRE20-like protein; tumour; tyrosinase; tyrosinase inhibitor-like protein; gene therapy; haematopoietic; illness; immunological disorder; neurodegenerative disorder; Alzheimer's disease; Parkinson's disease; immunomodulator; pharmacogenomic; haemostatic; human immunodeficiency virus; HIV; fertility disorder; neuroprotective; cytoskeletal; nontropic; anti-infertility; cancer; NOV-4b protein.

OS Homo sapiens.

FH	key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

FT	Peptide	1..22
FT	/label= signal_peptide	
FT	23..497	
FT	/note= "Human mature novel trypsin inhibitor-like protein NOV-4b"	
FT	195..206	
FT	/label= SCP_domain	

PN WO200162928-A2

PD 30-AUG-2001.

PF 26-FEB-2001; 2001WO-US06151.

PR. 25-FEB-2000; 2000US-0184951

PR 01-MAR-2000; 2000US-0185967

PR 27-APR-2000; 2000US-0199957

XX
XX
(CITIZEN -) CITIZEN CORP

XX
DT
VICEROY + CMM
Fernando de

XX WPB: 2001-582051/65.
DB

DR N-PSDB; AAD1//64.
XY

PT New Isolated ALVALZ
PT polypeptide for diag

Parkinson's disease
XX
XX

CS C:\MAIL L; Page 00, 10
XX

and their corresponding nucleic acid sequences. NOV-X collective

polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel

CC SREZ20 are polypeptides and NOV-X, NOV-Y, NOV-Z and NOV-X¹ are used to
CC identify a potential therapeutic agent that can modulate its activity
CC and can be used for treating a pathology related to aberrant expression
CC or aberrant physiological interactions of NOV-X, NOV-X or its DNA is
CC used to determine the presence or predisposition to a disease associated
CC with altered levels of NOV-X, NOV-X, its DNA and its antibody are used
CC to treat or prevent a pathology associated with NOV-X. The pathological
CC states that can be treated or prevented are haematopoietic, cancer,
CC immunological, tumour, neurodegenerative (e.g. Alzheimer's and
CC Parkinson's disease), human immunodeficiency virus (HIV) illness and
CC fertility disorders. NOV-X and its DNA are used in pharmacogenomics for
CC predictive medicine. NOV-X DNA is used in gene therapy. The present
CC sequence is human novel trypsin inhibitor-like protein, NOV-db.

Sequence 497 AA:

Query Match	99.4%	Score 2743;	DB 22;	Length 497;
Similarity	99.6%	Pred. No. 8.6e-227		
Matches 493; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

Qy	1	MSCVLGCVITPIGLIFLVCYSGSYLLPNPTILBEELLSTKOHNSHSHSVRAIRIPREDEELL	60
Db	1	mscvlgcvitpigliiflvcsqyilipnvtllbeellstkyamshsvrrairpredceell	60
Qy	61	MLHNKLRQVOPQASNMETWDELDEKSAAMASQCIENHPTSLVLSIGONLGAHMR	120
Db	61	mlhnlirgvgpqasnmeymtwddeleksaaawaqcienhptglilvsigqlgahwgr	120
Qy	121	YRSPGFHVOASWIDEYKDYTYTPRSECNPWCPRCSGPMCTHTYTQIYMAATTNKIGCAVNTC	180
Db	121	yrspgfhvqswydevkdytytyprsecnmpcprcsgpmcthtyqiymatlnkigcavntc	180
Qy	181	KRMVAGEWENAVYFVVCNITSPKGNMIGAPTKNGRPSCECPSPSYGSGCRNNLCYHEFTY	240

```

Db      181 fkmfhwgevenavayfvcnyspknwlgaeaykngtrpcscppsygscrmnlycreety 240
      |||
Qy      241 TPKEETDENMEVEETAPLPEENHVMLOPRVMPRTKPKTSANVNTMTQVVRCDTKMKDRCKG 300
      |||
Db      241 tlpkpetdemnevetapripeenhvwlqprvmpclpkpktsavnyntqvyrcdtkmkdrckg 300
      |||
Qy      301 STCRRYOCPAGCLNHHKAKIFGSLFEYSSSSICRAAIHIGLIDDKGGLVDITRNGKVPFFV 360
      |||
Db      301 stctnrycpagclnhkakiifgslfeyssssicraahyglliddkgylvdltrngkvpffv 360
      |||
Qy      361 KSEHGVOSLSKYKPPSSFMVSKVKVODLDQYTTVAQLCPFEKPATHCPRIHCPAHCKDE 420
      |||
Db      361 kserhgvqslskypssfmvskkvqldldqytlvaqlcpfekpathcprihpcahckde 420
      |||
Qy      421 PSYWAPFGNTNYADTSSICKTAVHAGVINSNCGVDVMPVDKRTYVGSILRNGVQSES 480
      |||
Db      421 psywarpfgntnyadtssicktavhagvinsnsgvdvmpvdkkkyvgsilrngvqses 480
      |||
Qy      481 LGTPRDGKAFRIFAVRQ 497
      |||
Db      481 lgtprdgkafrifavrq 497
      |||

RESULT 3
AAE10618
ID      AAE10618 standard; Protein; 498 AA.
AC      AAE10618;
DE      10-DEC-2001 (first entry)
XX      Human novel trypsin inhibitor-like protein, NOV-4d.
XX      Human; NOV-X protein; KIAA1233-like protein; STE20-like protein; tumour;
XX      trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
XX      immunological disorder; neurodegenerative disorder; Alzheimer's disease;
XX      Parkinson's disease; immunomodulatory; pharmacogenomic; hemostatic;
XX      human immunodeficiency virus; HIV; fertility disorder; neutroprotective;
XX      cytosolic; neutroptic; anti-fertility; cancer; NOV-4d protein.
XX      Homo sapiens.
XX      OS
XX      FH      Key      Location/Qualifiers
XX      FT      Domain      196..207
XX      FT      /label= SCP_domain
XX      PN      WO200162928-A2.
XX      PD      30-AUG-2001.
XX      PF      26-FEB-2001; 2001WO-US06151.
XX      PR      25-FEB-2000; 2000US-0184951.
XX      PR      28-FEB-2000; 2000US-0185448.
XX      PR      01-MAR-2000; 2000US-0185967.
XX      PR      18-APR-2000; 2000US-0197723.
XX      PR      27-APR-2000; 2000US-0199957.
XX      PR      23-FEB-2001; 2001US-0789390.
XX      PA      (CURA-) CUBAGEN CORP.
XX      PI      Vernet CAM, Fernandes E, Shinkets RA, Macdougall J, Spaderna SK;
XX      DR      N-PSDB; AAD17766.
XX      PT      WPI; 2001-582051/65.
XX      PT      New isolated KIAA1233-like, STE20-like, or trypsin inhibitor-like
XX      PT      polypeptide for diagnosing and treating pathological disorders, such as
XX      PT      Parkinson's disease and for use in pharmacogenomics -
XX      PS      Claim 1; Page 88; 189pp; English.
XX

```

```

CC      The invention relates to novel human polypeptides referred as NOV-X
CC      and their corresponding nucleic acid sequences. NOV-X collectively
CC      include NOV-1, NOV-2a and NOV-2b which are novel KIAA1233-like
CC      polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel
CC      STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e
CC      which are novel trypsin inhibitor-like polypeptides. NOV-X is used to
CC      identify a potential therapeutic agent that can modulate its activity
CC      and can be used for treating a pathology related to aberrant expression
CC      or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
CC      used to determine the presence or predisposition to a disease associated
CC      with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
CC      to treat or prevent a pathology associated with NOV-X. The pathological
CC      states that can be treated or prevented are haematopoietic, cancer,
CC      immunological, tumour, neurodegenerative (e.g. Alzheimer's and
CC      Parkinson's disease), human immunodeficiency virus (HIV) illness and
CC      fertility disorders. NOV-X and its DNA are used in pharmacogenomics for
CC      predictive medicine. NOV-X DNA is used in gene therapy. The present
CC      sequence is human novel trypsin inhibitor-like protein, NOV-4d.
XX      SQ      Sequence      498 AA;

```

```

Query Match      99.2%; Score 2736.5; DB 22; Length 498;
Best Local Similarity 99.6%; Pred. No. 3.1e-226;
Matches 496; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```

```

Qy      1 MSCVLAGVIVPLGLFLVNGSOGYLLPNVTLLELLSKYQHNSHRRARIPREKEITL 60
      |||
Db      1 mscvlagvivplgllflvngsgyillpnvtlleellskymeshrraripredkeil 60
      |||
Qy      61 MLHNKLRGQVOPQASNNEMYTWDELEKSAAMASQCIWEHGPTSLVYSIGQNGAHWG- 119
      |||
Db      61 mlhnklrgqvopqasnnemytwdelaksaasqciwhegptslvysigngahwgr 120
      |||
Qy      120 RYRSPGFVQSVWDEVDYRTPYSECPWCPRECSGPMCHRYQIYMATTKIGCAVNT 179
      |||
Db      121 ryrspgfhwsvwdevdyrtpysecpwcpresgpmchryqiywattkigcavnt 180
      |||
Qy      180 CRKMTWGEWENAVFYVCNYSPPKGNWIGEAIFYKNGRPSCEPSYSGSCRNNLCYREET 239
      |||
Db      181 crkmtwgevenavayfvcnyspknwlgaeaykngtrpcscppsygscrmnlycreet 240
      |||
Qy      240 YTPKEETDENMEVEETAPLPEENHVMLOPRVMPRTKPKTSANVNTMTQVVRCDTKMKDRCK 299
      |||
Db      241 ytpkpetdemnevetapripeenhvwlqprvmpclpkpktsavnyntqvyrcdtkmkdrck 300
      |||
Qy      300 GSTCNRYOCPAGCLNHHKAKIFGSLFEYSSSSICRAAIHIGLIDDKGGLVDITRNGKVPFF 359
      |||
Db      301 gstcnrycpagclnhkakiifgslfeyssssicraahyglliddkgylvdltrngkvpff 360
      |||
Qy      360 VKSEHGVOSLSKYKPPSSFMVSKVKVODLDQYTTVAQLCPFEKPATHCPRIHCPAHCKD 419
      |||
Db      361 vkserhgvqslskypssfmvskkvqldldqytlvaqlcpfekpathcprihpcahckd 420
      |||
Qy      420 EPSYWAPFGNTNYADTSSICKTAVHAGVINSNCGVDVMPVDKRTYVGSILRNGVQSES 479
      |||
Db      421 epsywarpfgntnyadtssicktavhagvinsnsgvdvmpvdkkkyvgsilrngvqse 480
      |||
Qy      480 SLGTPRDGKAFRIFAVRQ 497
      |||
Db      481 slgtprdgkafrifavrq 498
      |||

RESULT 4
AAE10619
ID      AAE10619 standard; Protein; 501 AA.
AC      AAE10619;
DE      10-DEC-2001 (first entry)
XX      Human novel trypsin inhibitor-like protein, NOV-4e.
XX

```

KM Human: NOV-X protein; KIAA1233-like protein; STE20-like protein; tumour;
 KM trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
 KM immunological disorder; neurodegenerative disorder; Alzheimer's disease;
 KM Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
 KM human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
 KM cytosolic; neotropic; anti-fertility; cancer; NOV-4e protein.
 OS Homo sapiens.
 XX
 XX
 PH Key Location/Qualifiers
 FT Peptide 1..22
 FT /label= Signal_peptide
 FT 23..501
 FT /note= "Human mature novel trypsin inhibitor-like
 protein, NOV-4e"
 FT
 PD WO200162928-A2.
 PD 30-AUG-2001.
 PD 26-FEB-2001; 2001WO-US06151.
 PD
 XX 25-FEB-2000; 2000US-0184951.
 PR 28-FEB-2000; 2000US-0185348.
 PR 01-MAR-2000; 2000US-0185967.
 PR 18-APR-2000; 2000US-0197723.
 PR 27-APR-2000; 2000US-0199957.
 PR 23-FEB-2001; 2001US-0789390.
 PR
 XX (CURA-) CURAGEN CORP.
 PA
 XX Vernet CAM, Fernandes E, Shinkets RA, Macdougall J, Spaderna SK;
 PI
 XX WPI; 2001-582051/65.
 DR N-PSDB; AAD17767.
 DR
 XX New isolated KIAA1233-like, STE20-like, or trypsin inhibitor-like
 PT polypeptide for diagnosing and treating pathological disorders, such as
 PT Parkinson's disease and for use in pharmacogenomics -
 PS
 PS Claim 1; Page 91; 189pp; English.
 XX
 XX The invention relates to novel human polypeptides referred as NOV-X
 CC and their corresponding nucleic acid sequences. NOV-X collectively
 CC include NOV-1, NOV-2a and NOV-2b which are novel KIAA1233-like
 CC polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel
 CC STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e
 CC which are novel trypsin inhibitor-like polypeptides. NOV-X is used to
 CC identify a potential therapeutic agent that can modulate its activity
 CC and can be used for treating a pathology related to aberrant expression
 CC or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
 CC used to determine the presence or predisposition to a disease associated
 CC with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
 CC to treat or prevent a pathology associated with NOV-X. The pathological
 CC states that can be treated or prevented are haematopoietic, cancer,
 CC immunological, tumour, neurodegenerative (e.g. Alzheimer's and
 CC Parkinson's disease), human immunodeficiency virus (HIV) illness and
 CC fertility disorders. NOV-X and its DNA are used in pharmacogenomics for
 CC predictive medicine. NOV-X DNA is used in gene therapy. The present
 CC sequence is human novel trypsin inhibitor-like protein, NOV-4e.
 CC
 XX Sequence 501 AA:
 SQ

Query Match 98.6%; Score 2721; DB 22; Length 501;
 Best Local Similarity 98.8%; Pred. No. 6,7e-225;
 Matches 495; Conservative 0; Mismatches 2; Indels 4; Gaps 2;

QY 1 MSCVIGVPIGLILVCGSGYLLPNVTLLEBLISKYQNHESHRRRAIPRDKKEIL 60
 DB 1 macvlgvpiplgllflivrgsgyllpnvtlleeliskyqheshrrraiprdekell 60
 QY 61 MLHNKLRGVQVQASMEFWMTWDELKSAAMAASQCIWHEGPTSLVLSGNGAIMGK 120

DB 61 mlhnklrgvqpqasmeymtwdelksaawasqciwhegptllvlsigqlgahvgr 120
 QY 121 YRSRGFVQSVYDVKQYTYPRPSECNMPCEKRGPMCTHY---TQVMAATNKIGAV 177
 DB 121 yrspgflvqsvydvkdylyprsecnmpceksqgmcthytqvtlwtalnkgav 180
 QY 178 NTCRKMTVWGEWENAVFYVGNYSRK-GNWIGEARPYKNGRPSCEKPSYSGSCRNMLCYR 236
 DB 181 ntcrkmtvgewenavfyvgnysrknwigearykngrpsceppysgscrnmlcyr 240
 QY 237 EETYPKPEIDENNEVEETAPLPEENHWLQPRWAPPTPKTSAVNYMTQVRCOTKAKD 296
 DB 241 eetylprkpeidennvevetaplppeenhwlpdwmapptpkktisavnymtqvrrcdtkmkd 300
 QY 297 RCKGTCNRVOCPCAGCLNKKAKIFGSLFEYSSSSICRAIHGILDDGGGLVDITRNKGV 356
 DB 301 rckgstcnryqpcagclnnkakilfgslfeysssicraahgildgglvdltrngkv 360
 QY 357 PFFVKSERHGVQSLSKKRPSSSFVMSKVKVQDLDCTTVAQLCPREKATPHCRHCPAH 416
 DB 361 pffvkserrhgvqslskkrpsssfvmskvkvqdlldcvtvaqlcprekathcprhchpah 420
 QY 417 CKDEPSTWAPYFGTNTVADTSSICTAVHAGVISMESGDDVMDKKTGVSLRNGV 476
 DB 421 ckdepswapyfgtnlyadtsaiclavhagvisnesgddvmdpdkkklvyslrlngv 480
 QY 477 QSESLGTPRDGKAFFRIFAVRQ 497
 DB 481 qseislgtprdgkaffrifavrq 501

RESULT 5
 AAE10617
 ID AAE10617 standard; Protein; 477 AA.
 XX
 AC AAE10617;
 DT 10-DEC-2001 (first entry)
 XX
 DE Human novel trypsin inhibitor-like protein, NOV-4c.
 XX
 KM Human: NOV-X protein; KIAA1233-like protein; STE20-like protein; tumour;
 KM trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
 KM immunological disorder; neurodegenerative disorder; Alzheimer's disease;
 KM Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
 KM human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
 KM cytosolic; neotropic; anti-fertility; cancer; NOV-4c protein.
 OS Homo sapiens.
 XX
 XX
 PH Key Location/Qualifiers
 FT Peptide 1..22
 FT /label= Signal_peptide
 FT 23..477
 FT /note= "Human mature novel trypsin inhibitor-like
 protein, NOV-4c"
 FT Domain 81..92
 FT /label= SCP_domain
 FT
 PD WO200162928-A2.
 PD 30-AUG-2001.
 PD 26-FEB-2001; 2001WO-US06151.
 PD
 XX 25-FEB-2000; 2000US-0184951.
 PR 28-FEB-2000; 2000US-0185348.
 PR 01-MAR-2000; 2000US-0185967.
 PR 18-APR-2000; 2000US-0197723.
 PR 27-APR-2000; 2000US-0199957.
 PR 23-FEB-2001; 2001US-0789390.
 PR
 XX

PA (CURA-) CURAGEN CORP.
 XX Vernet CAM, Fernandes E, Shinkets RA, MacDougall J, Spaderna SK;
 PI MPI: 2001-582051/65.
 DR N-PSDB; AAD17765.
 XX
 PT New isolated KIAA1233-like, STE20-like, or trypsin inhibitor-like
 PT polypeptide for diagnosing and treating pathological disorders, such as
 PT Parkinson's disease and for use in pharmacogenomics -
 XX
 PS Claim 1: Page 85-86; 189pp; English.
 XX
 CC The invention relates to novel human polypeptides referred as NOV-X
 CC and their corresponding nucleic acid sequences. NOV-X collectively
 CC include NOV-1, NOV-2a and NOV-2b which are novel KIAA1233-like
 CC polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel
 CC STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e
 CC which are novel trypsin inhibitor-like polypeptides. NOV-X is used to
 CC identify a potential therapeutic agent that can modulate its activity
 CC and can be used for treating a pathology related to aberrant expression
 CC or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
 CC used to determine the presence or predisposition to a disease associated
 CC with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
 CC to treat or prevent a pathology associated with NOV-X. The pathological
 CC states that can be treated or prevented are haematopoietic, cancer,
 CC immunological, tumour, neurodegenerative (e.g. Alzheimer's and
 CC Parkinson's disease), human immunodeficiency virus (HIV) illness and
 CC fertility disorders. NOV-X and its DNA are used in pharmacogenomics for
 CC predictive medicine. NOV-X DNA is used in gene therapy. The present
 CC sequence is human novel trypsin inhibitor-like protein, NOV-4c.
 XX
 SQ Sequence 477 AA:

Query Match 94.1%; Score 2595; DB 22; Length 477;
 Best Local Similarity 98.9%; Pred. No. 4.1e-214;
 Matches 463; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSCVGVGVPFLGLFLFLYCSGGYLLPNVTLLELLSKYOHNSHVKRAIPREDKEEIL 60
 DB 1 mscvgyvfpflglflflvcsggyllpnvlllellskymshsvrraiaedeeell 60
 QY 61 MLHNKRGQVOPQASNMETWDELEKSAAMASQCIHEHPTSLVSTIGNLGAHGR 120
 DB 61 mlhnklrgvqpqaasmeymtdelksaaasgcivehptslvstignlgahgr 120
 QY 121 YRSPGFHQSMTDEKDYTPPSCNPMCPBRCGPMCTHYQTIVMATNTKIGCAVNTC 180
 DB 121 yrspgfhwswydevkdytppscnmpcbrcgpmcthyqtivmatntkigcavntc 180
 QY 181 RKMTVGEVMEENAVFYCNYSRKGWIGEPYKNGRPSCEPSSYSGSCRNLCTREETY 240
 DB 181 rkmtvgevmenavfyvncysrkgwigeapkyngprscsqpsysgscrnlctreety 240
 QY 241 TPKEETDAMNEVEFAPIDPEENHVMLOPRVMPRTKRTKSANVMYQVVRCDTKMDRCKG 300
 DB 241 tpkeetdamevefapideenhvmloprvmprtkrktksanvmyqvvrctdkmdrckg 300
 QY 301 STCNRYOCPACCLNHNKAKIFGSLYEESSTICRAIHGILDDKGLVDITRNKAVPEFV 360
 DB 301 stcnryocpaccclnhnkaki fgsllyeesssticraihgilddkgylvditrnkavpfv 360
 QY 361 KSEHGVQSLSKYRPSSEFMYSKVQVODLDCYTTVAQCPFEKPAHCPRIHCPHNCDE 420
 DB 361 ksehgvqslskyrpssefmyskvqvodldcyyttvaqcpfekpahcprhchphncde 420
 QY 421 PSYAPFEGTNTIVADTSSICTAVHAGVISMESGDDVMPVKKKTY 468
 DB 421 psyapfegntivadtssictavhagvismesgddvmpvkkkty 468
 RESULT 6

AAE10615
 ID AAE10615 standard; Protein; 383 AA.
 XX
 AC AAE10615;
 XX
 DT 10-DEC-2001 (first entry)
 XX
 DE Human novel trypsin inhibitor-like protein, NOV-4a.
 XX
 KW Human; NOV-X protein; KIAA1233-like protein; STE20-like protein; tumour;
 KW trypsin inhibitor-like protein; gene therapy; haematopoietic; illness;
 KW immunological disorder; neurodegenerative disorder; Alzheimer's disease;
 KW Parkinson's disease; immunomodulatory; pharmacogenomic; haemostatic;
 KW human immunodeficiency virus; HIV; fertility disorder; neuroprotective;
 KW cytosolic; neurotropic; anti-fertility; cancer; NOV-4a protein.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FT Domain
 FT /label= SCP-domain
 XX
 PN W0200162928-A2.
 XX
 PD 30-AUG-2001.
 XX
 PF 26-FEB-2001; 2001MO-US06151.
 XX
 PR 25-FEB-2000; 2000US-0184951.
 PR 28-FEB-2000; 2000US-0185548.
 PR 01-MAR-2000; 2000US-0185967.
 PR 18-APR-2000; 2000US-0197723.
 PR 27-APR-2000; 2000US-0199957.
 PR 23-FEB-2001; 2001US-0789390.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Vernet CAM, Fernandes E, Shinkets RA, MacDougall J, Spaderna SK;
 DR MPI: 2001-582051/65.
 DR N-PSDB; AAD17763.
 XX
 PT New isolated KIAA1233-like, STE20-like, or trypsin inhibitor-like
 PT polypeptide for diagnosing and treating pathological disorders, such as
 PT Parkinson's disease and for use in pharmacogenomics -
 XX
 PS Claim 1: Page 81; 189pp; English.
 XX
 CC The invention relates to novel human polypeptides referred as NOV-X
 CC and their corresponding nucleic acid sequences. NOV-X collectively
 CC include NOV-1, NOV-2a and NOV-2b which are novel KIAA1233-like
 CC polypeptides, NOV-3a, NOV-3b, NOV-3c and NOV-3d which are novel
 CC STE20-like polypeptides and NOV-4a, NOV-4b, NOV-4c, NOV-4d and NOV-4e
 CC which are novel trypsin inhibitor-like polypeptides. NOV-X is used to
 CC identify a potential therapeutic agent that can modulate its activity
 CC and can be used for treating a pathology related to aberrant expression
 CC or aberrant physiological interactions of NOV-X. NOV-X or its DNA is
 CC used to determine the presence or predisposition to a disease associated
 CC with altered levels of NOV-X. NOV-X, its DNA and its antibody are used
 CC to treat or prevent a pathology associated with NOV-X. The pathological
 CC states that can be treated or prevented are haematopoietic, cancer,
 CC immunological, tumour, neurodegenerative (e.g. Alzheimer's and
 CC Parkinson's disease), human immunodeficiency virus (HIV) illness and
 CC fertility disorders. NOV-X and its DNA are used in pharmacogenomics for
 CC predictive medicine. NOV-X DNA is used in gene therapy. The present
 CC sequence is human novel trypsin inhibitor-like protein, NOV-4a.
 XX
 SQ Sequence 383 AA:

Query Match 77.6%; Score 2140; DB 22; Length 383;
 Best Local Similarity 99.5%; Pred. No. 3.4e-175;
 Matches 379; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

SEA ID NO: 2
AC NO: AAY41738, Database: A. Geneseg-032802

```

OY 117 HMGRRSPGFHVSQWDEVDKYTPYPSPSCNPMCPERCSCPMCTHTQIWMATTNKGCA 176
DB 3 nwgrrspgfhwgswvdekytppspscnmpcpcscgmchtyqjwattknkgca 62
OY 177 VNTCRKMTWGEWENAVYFVCNYSRKGWITGAPYKNGRPSRCSGSGCNNTCYR 236
DB 63 vntcrkmtwgewenavyfvcnysrkgwvlgapyngrpsrscsgsgcnntcyr 122
OY 237 EETVKKPPEDEKNEVETAPRPEENHWLQPRVMPKPKKTSANVMTOVVRCDTKMKD 296
DB 123 eetvkkppetdemnevetapreennhwlpvrmprkpkktasvnmtyvrcdtkmkd 182
OY 297 RCKGSTCNRYQCPAGCLNHRKAKTFGLFSSSSSTICRAAIHYGIILDRKGGLVDITRNKV 356
DB 183 rckgstcnryqcpagclnhkakifglfysesssicraahygilldrkglvdltrngkv 242
OY 357 PPRVKKERRHGVOSTLSKYKPSFVSVKVVODLDCYTTVAOLCPFEKPRTHCPRHCPAH 416
DB 243 pprvkkerrhgvostlskykpsfsvsvkvvoddldcyttvaolcpfekprthcprhchpah 302
OY 417 CKDEPSYMAVFEGTNIYADNSSLCKTAVHAGVISNESGGDVMPVDKKTYYVGSILNGV 476
DB 303 ckdepsymavfegtntiyadnsslcktavhagvisnesggdvmpvdkktyvgsilrngv 362
OY 477 QSESLSGTPRDKAPRIRAVNQ 497
DB 363 qseslsgtpdrgkafiriravrq 383

```

RESULT 7

AAB93979 standard; Protein: 371 AA.

26-JUN-2001 (first entry)

Human protein sequence SEQ ID NO:14059.

Human; primer: detection; diagnosis; antisense therapy; gene therapy.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-0116126.

29-JUL-1999; 99JP-0248036.

27-AUG-1999; 99JP-0300253.

11-JAN-2000; 2000JP-0118776.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

WPI: 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602

full-length cDNAs defined in the specification, and for the detection

and/or diagnosis of the abnormality of the proteins encoded by the

full-length cDNAs -

Claim 8; SEQ ID 14059; 2537bp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602

full-length cDNAs defined in the specification. Where a primer set

comprises: (a) an oligo-dT primer and an oligonucleotide complementary

to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 371 AA;

Query Match 75.1%; Score 2071; DB 22; Length 371;
Best Local Similarity 99.7%; Pred. No. 2.7e-169;
Matches 369; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 MSCVLGAVTPLGLFLVCGSGYLTPVNTLLELLLSKYQHNSHRYRRAIRPDKKEIL 60
DB 1 mscvlgavtplglflvcgsqgyltpvntllelllskyqhnsrvtirralpredkeell 60
OY 61 MLHNKLRGQVQPOASNMETWDELEKSAANAASOCITWEGPSTSLVSTQNGAHRGR 120
DB 61 mlhnklrgqvqpasnmetytwdeleksaanaasocitwegpstsllvsgnlgahwgr 120
OY 121 YRSPGFHVOSWYEVKRYTPYPSECNPMCPERCSCPMCTHTQIWMATTNKGCAVNTC 180
DB 121 yrspgfhwswyevkrytppspscnmpcpcscgmchtyqjwattknkgcavntc 180
OY 181 RKMTWGEWENAVYFVCNYSRKGWITGAPYKNGRPSRCSGSGCNNTCYRETY 240
DB 181 rkmtwgewenavyfvcnysrkgwvlgapyngrpsrscsgsgcnntcyrcty 240
OY 241 TPKPETDEMNEVETAPRPEENHWLQPRVMPKPKKTSANVMTOVVRCDTKMKDCKG 300
DB 241 tpkpetdemnevetapreennhwlpvrmprkpkktasvnmtyvrcdtkmkdckg 300
OY 301 STCNRYQCPAGCLNHRKAKTFGLFSSSSSTICRAAIHYGIILDRKGGLVDITRNKV 360
DB 301 stcnryqcpagclnhkakifglfysesssicraahygilldrkglvdltrngkv 360
OY 361 KSERHGVOSTLSKYKPSFVSVKVVODLDCYTTVAOLCPFEKPRTHCPRHCPAH 416
DB 361 kserhgvostlskykpsfsvsvkvvoddldcyttvaolcpfekprthcprhchpah 416

```

RESULT 8

AAV41738 standard; Protein: 500 AA.

AAV41738;

07-DEC-1999 (first entry)

Human PRO541 protein sequence.

Human; PRO: EST; expressed sequence tag; PCR primer; hybridisation;

probe; blood coagulation disorder; cancer; cellular adhesion disorder;

secreted protein; transmembrane protein.

Homo sapiens.

WO9946281-A2.

XX	16-SEP-1999.	99MO-0505028
XX	08-MAR-1999.	
PF		
XX		
PR	10-MAR-1998.	98US-0077450
PR	11-MAR-1998.	98US-0077632
PR	11-MAR-1998.	98US-0077641
PR	11-MAR-1998.	98US-0077649
PR	12-MAR-1998.	98US-0077791
PR	13-MAR-1998.	98US-0078004
PR	17-MAR-1998.	98US-0040220
PR	20-MAR-1998.	98US-0078886
PR	20-MAR-1998.	98US-0078910
PR	20-MAR-1998.	98US-0078936
PR	25-MAR-1998.	98US-0079294
PR	26-MAR-1998.	98US-0079656
PR	27-MAR-1998.	98US-0079663
PR	27-MAR-1998.	98US-0079664
PR	27-MAR-1998.	98US-0079728
PR	27-MAR-1998.	98US-0079786
PR	30-MAR-1998.	98US-0079920
PR	30-MAR-1998.	98US-0079923
PR	31-MAR-1998.	98US-0080105
PR	31-MAR-1998.	98US-0080107
PR	31-MAR-1998.	98US-0080165
PR	31-MAR-1998.	98US-0080171
PR	01-APR-1998.	98US-0080327
PR	01-APR-1998.	98US-0080328
PR	01-APR-1998.	98US-0080334
PR	01-APR-1998.	98US-0080334
PR	08-APR-1998.	98US-0081049
PR	08-APR-1998.	98US-0081070
PR	08-APR-1998.	98US-0081071
PR	09-APR-1998.	98US-0081193
PR	09-APR-1998.	98US-0081205
PR	15-APR-1998.	98US-0081129
PR	15-APR-1998.	98US-0081817
PR	15-APR-1998.	98US-0081838
PR	15-APR-1998.	98US-0081952
PR	21-APR-1998.	98US-0082568
PR	21-APR-1998.	98US-0082569
PR	22-APR-1998.	98US-0082700
PR	22-APR-1998.	98US-0083704
PR	23-APR-1998.	98US-0082804
PR	23-APR-1998.	98US-0082867
PR	27-APR-1998.	98US-0083796
PR	28-APR-1998.	98US-0083736
PR	29-APR-1998.	98US-0083322
PR	29-APR-1998.	98US-0083495
PR	29-APR-1998.	98US-0083496
PR	29-APR-1998.	98US-0083499
PR	29-APR-1998.	98US-0083500
PR	29-APR-1998.	98US-0083543
PR	29-APR-1998.	98US-0083554
PR	29-APR-1998.	98US-0083558
PR	30-APR-1998.	98US-0083559
PR	30-APR-1998.	98US-0083742
PR	05-MAY-1998.	98US-0084366
PR	06-MAY-1998.	98US-0084414
PR	06-MAY-1998.	98US-0084441
PR	07-MAY-1998.	98US-0084598
PR	07-MAY-1998.	98US-0084600
PR	07-MAY-1998.	98US-0084627
PR	07-MAY-1998.	98US-0084637
PR	07-MAY-1998.	98US-0084639
PR	07-MAY-1998.	98US-0084640
PR	07-MAY-1998.	98US-0084643
PR	13-MAY-1998.	98US-0085323

(GETH) GENENTECH INC.
Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
WPI; 1999-551358/46.
N-PSDB; AA234199.
New secreted and transmembrane polypeptides and their polynucleotides
useful for treating blood coagulation disorders, cancers and cellular
adhesion disorders -
Claim 12; Fig 147; 530pp; English.

Query Match	59.58;	Score 1640.5;	DB 20;	Length 500;
Best Local Similarity	59.58;	Pred. No. 3.7e-13;		
Matches 292;	Conservative 74;	Mismatches 112;	Indels 13;	Gaps 7

[illegible]

```

OY 305 RYOCPCACGLNKKAKIFGSLFYESSSSICRAAIHYGLIDDKGLVDITRNGKVPFVKSR 364
DB 310 rycpcagclskakvlgsvhyemqssicraahyglindggwvdltrgrkhyfiksnt 369
OY 365 HGVOSLSKYKRPSSFMWSKVYQDLDCYTYVAQLCPFEKPRATHCRPHACPKDEPSYW 424
DB 370 ngqtlgkygansftvskvgtvgtcvtelqclpfkpkashpcrvygcprncmqanphy 429
OY 425 APVFGTNIYADTSSICRTAVHAGVINSNCGDVPDVPDKKRTYVGLRNGVQSSSLGTP 484
DB 430 arvlgtrvysldssicraavhagvvrnh-ggyvdympdkrtktylasfnglfseslgnp 488
OY 485 RDGKAFRIFAV 495
DB 489 pggkaftrfav 499

RESULT 9
ID AAB44294 standard; Protein: 500 AA.
AAB44294:
AC AAB44294:
XX
DT 08-FEB-2001 (first entry)
XX
DE Human PRO541 (UNQ342) protein sequence SEQ ID NO:363.
XX
KW Human: secreted protein; transmembrane protein; PRO; EST; cytosolic;
XX expressed sequence tag; detection; cancer.
XX
OS Homo sapiens.
XX
PN WO200053756-A2.
XX
PD 14-SEP-2000.
XX
PF 18-FEB-2000; 2000WO-US04341.
XX
PR 08-MAR-1999; 99WO-US05028.
XX 12-MAR-1999; 99US-0123957.
XX 29-MAR-1999; 99US-0126773.
XX 21-APR-1999; 99US-0130232.
XX 28-APR-1999; 99US-0131445.
XX 14-MAY-1999; 99US-0134287.
XX 23-JUN-1999; 99US-0141037.
XX 26-JUL-1999; 99US-0145698.
XX 29-OCT-1999; 99US-0162506.
XX 30-NOV-1999; 99WO-US28313.
XX 02-DEC-1999; 99WO-US28551.
XX 02-DEC-1999; 99WO-US28565.
XX 16-DEC-1999; 99WO-US30095.
XX 30-DEC-1999; 99WO-US31243.
XX 30-DEC-1999; 99WO-US31274.
XX 05-JAN-2000; 2000WO-US00219.
XX 06-JAN-2000; 2000WO-US00277.
XX 06-JAN-2000; 2000WO-US00376.
XX
PA (GETH ) GENENTECH INC.
XX
PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerlitsen ME;
PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
PI Kijavini ID, Kuo SS, Napier MA, Pan J, Paoletti NF, Roy MA;
PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
XX
XX WPI: 2000-611443/58.
XX N-PSDB: AAC78550.
XX
XX Novel PRO polypeptides and polynucleotides used in detection methods,
XX to target bioactive molecules to specific cells, and to modulate
XX cellular activities -
XX
XX Claim 12; Fig 147; 636pp; English.

```

```

XX AAC78458 to AAC78599 represent polynucleotide and EST (expressed
CC sequence tag) sequences which encode secreted or transmembrane PRO
CC polypeptides. The PRO polynucleotides and polypeptides have cytosolic
CC activity. The polynucleotides and polypeptides can be used for detecting
CC the presence of PRO polypeptides in samples, for linking bioactive
CC molecules to cells and for modulating biological activities of cells,
CC using the polypeptides for specific targeting. The polypeptide targeting
CC can be used to kill the target cells, e.g. for the treatment of cancers.
CC The polypeptide pairs provide specific targeting of bioactive molecules
CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
CC the isolation of the PRO polynucleotide sequences.
XX
XX Sequence 500 AA;
XX
XX Query Match 59.5%; Score 1640.5; DB 21; Length 500;
XX Best Local Similarity 59.5%; Pred. No. 3,76-132;
XX Matches 292; Conservative 74; Mismatches 112; Indels 13; Gaps 7;
XX
OY 13 LRLFLVCGSGGYLPNTVLEELLSKYQHN-----SHSRVRRATIPREDKEILMLNKLK 67
DB 14 vlfmarapamvvpnatlleklekyndedegwjakrgkratldndmgslldlnklr 73
OY 68 GQVQPOASNMETWTDDELEKSAAMASQCTIHEGPTSLVSGONLCAHMGRTSPGFH 127
DB 74 sqvyptasmeymtwdvelersaeswesciwehgnpsalipsiqnlgahwyrypctfh 133
OY 128 VOSWYDEVKDVTYPPSPBCNPMCPERCSPMCTHTYTOIWMATTNKIGAVNCRKMTWG 187
DB 134 vqswydeykdtsyphecnmpcfrcspgpccthyqvvatsnrlgcalnchmmlwg 193
OY 188 EWMENAVYFVNCNPSPKNGWIGAPYKNGRPGSCSPSYGSGCRNMLCYRE--EYTPPE 245
DB 194 qfwpkavylvncnpskgwghapkykhrpsacpsfsggorenlyksgsdrlyp-pr 252
OY 246 TDEANVEYETAPIP-EENNVMYQPRMRTKPKTSAAVVMYMOYVRCDKMRDRGSGTCN 304
DB 253 eecnelelrgsqvhdltv--rtstdssrnevlsa-qmqslyacevrlldqcgylton 309
OY 305 RYOCPCACGLNKKAKIFGSLFYESSSSICRAAIHYGLIDDKGLVDITRNGKVPFVKSR 364
DB 310 rycpcagclskakvlgsvhyemqssicraahyglindggwvdltrgrkhyfiksnt 369
OY 365 HGVOSLSKYKRPSSFMWSKVYQDLDCYTYVAQLCPFEKPRATHCRPHACPKDEPSYW 424
DB 370 ngqtlgkygansftvskvgtvgtcvtelqclpfkpkashpcrvygcprncmqanphy 429
OY 425 APVFGTNIYADTSSICRTAVHAGVINSNCGDVPDVPDKKRTYVGLRNGVQSSSLGTP 484
DB 430 arvlgtrvysldssicraavhagvvrnh-ggyvdympdkrtktylasfnglfseslgnp 488
OY 485 RDGKAFRIFAV 495
DB 489 pggkaftrfav 499

RESULT 10
ID AAU29058 standard; Protein: 500 AA.
AAU29058:
XX
XX AAU29058:
XX
DT 18-DEC-2001 (first entry)
XX
DE Human PRO polypeptide sequence #35.
XX
KW PRO polypeptide; mammary; tumour; cancer; human; cattle; horse; sheep;
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
XX
XX Homo sapiens.

```